Cloud Cover Predictions Diagnosed From Global Numerical Weather Prediction Model Forecasts

Donald C. Norquist
Donald L. Aiken
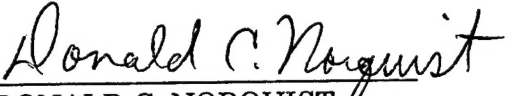Daniel A. DeBenedictus

7 February 1997

DTIC QUALITY INSPECTED 2

PHILLIPS LABORATORY
Directorate of Geophysics
AIR FORCE MATERIEL COMMAND
HANSCOM AIR FORCE BASE, MA 01731-3010

19971124 026

"This technical report has been reviewed and is approved for publication."

Donald C. Norquist

DONALD C NORQUIST
Acting Chief, Forecast Applications Branch
Optical Effects Division

Donald A. Chisholm

DONALD A/CHISHOLM
Acting Director
Optical Effects Division

This document has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Information Service (NTIS).

Qualified requesters may obtain additional copies from the Defense Technical Information Center (DTIC). All others should apply to the NTIS.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE 7 February 1997	3. REPORT TYPE AND DATES COVERED Final Report Oct 93-Sep 95

4. TITLE AND SUBTITLE Cloud Cover Predictions Diagnosed From Global Numerical Weather Prediction Model Forecasts	5. FUNDING NUMBERS PE 62101F PR 6670 TA GT
6. AUTHOR(S) Norquist, Donald C., Donald L. Aiken, and Daniel A. DeBenedictus	WU 35

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Phillips Laboratory (GPAB) 29 Randolph Road Hanscom Air Force Base, Massachusetts 01731-3010	8. PERFORMING ORGANIZATION REPORT NUMBER PL-TR-97-2015 ERP, No.1203

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We developed statistical relationships between fractional cloud cover and a collection of variables drawn from forecast fields from a global numerical weather prediction model. These relationships were then applied to later forecasts from the same weather prediction model to diagnose the cloud cover corresponding to the forecast states. The U.S. Air Force RTNEPH cloud analysis datasets in the Northern Hemisphere for January and July 1991 were used to represent cloud cover in separate winter and summer statistical relationship developments. Forecasts from the Phillips Laboratory Global Spectral Model at times corresponding to the cloud analyses were used to provide the weather model predictors. We used multiple linear regression and a hybrid regression estimation of event possibilities/multiple linear regression to develop the statistical relationships in the 5 percent cloud amount categories from 0 to 100 percent. We used multiple discriminant analysis with six cloud amount categories covering the range from 0 to 100 percent. Separate relationships were developed over 10-day periods for low, middle, and high cloud decks, and total cloud. The relationships developed were also distinct for different forecast durations. All relationships were applied to forecasts initialized on the day following the 10-day development period. Our results showed that multiple linear regression produced forecast diagnoses of cloud amount that were slightly better than the other methods in root-mean-square error. This method also modified the frequency distribution of cloud cover from that of the analysis. The hybrid method improved upon the frequency distribution of clear and overcast, but worsened the near-clear and near-overcast categories. Multiple discriminant analysis produced cloud cover diagnoses with the best combination of root-mean-square skill and preservation of the frequency distribution. MDA also produces an estimate of the probability of cloud amount category selected being the correct one.

14. SUBJECT TERMS Numerical weather prediction Model output statistics Cloud forecasting Multiple discriminant analysis	15. NUMBER OF PAGES 150
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR

Contents

Illustrations

Tables

Acknowledgment

Cloud Cover Predictions Diagnosed from Global Numerical Weather Prediction Model Forecasts

1. INTRODUCTION

The goal of this study was to develop and demonstrate methods to diagnose fractional cloud cover (hereafter, the term "cloud amount" will be used interchangeably with fractional cloud cover) from non-cloud global numerical weather prediction (NWP) model forecasts. The final report of the previous study[1] (hereafter, CldAmt94) describes the development and testing of a baseline statistical algorithm for relating analyses of cloud amount to non-cloud NWP forecasts. In the present study, we sought measurable improvement in cloud amount predictions diagnosed from non-cloud NWP forecasts (hereafter, cloud amount diagnoses) through improved cloud cover specifications, NWP forecasts, and statistical methods relating the two. Ultimately, our goal is to determine and demonstrate greatest skill level possible in large-scale NWP model-based forecast cloud amount diagnosis.

In its introduction, CldAmt94 includes a discussion of US Air Force requirements, and current capabilities, of cloud amount analyses and forecasts

produced operationally by the Air Force Global Weather Center (AFGWC). The current AFGWC operational cloud amount analysis model is called RTNEPH.[2,3] The current large-scale (100-200 km grid spacing) cloud amount forecast model run operationally at AFGWC is called 5LAYER.[4] The intent of this series of studies was to develop and demonstrate cloud forecast methods that might be candidates to provide cloud amount forecasts operationally in the future. Thus, the candidate scheme would have to demonstrate skill and usability advantages over the 5LAYER model to be considered a potential replacement for 5LAYER.

Improved cloud amount forecasts are of importance to the US Air Force in several ways. First, the warfighter benefits directly in terms of a more accurate knowledge of cloud cover for air-to-ground target acquisition probability. Second, air-to-air operations requiring line-of-sight (such as in-flight refueling) can be better planned with a better prediction of future cloud distribution. Thirdly, space re-entry can often be impacted by clouds, and an assessment of the adequacy of the planned re-entry location can be better obtained using improved cloud amount predictions. Fourth, air-to-ground reconnaissance flights can be better planned with a more accurate knowledge of future cloud obscuration of the ground from aircraft. Finally, aircraft traffic limitations due to visibility and ceiling restrictions at Air Force bases can be anticipated with greater accuracy if better cloud amount forecasts are available. For these reasons and more, accurate cloud amount forecasts remain a high priority requirement by the U.S. Air Force.

Cloud cover prediction methods can be classified into three general categories: diagnostic, trajectory, and prognostic. A diagnostic method uses available NWP model forecasts that do not include cloud amount as a predicted quantity, simultaneous cloud amount observations or analyses, and a statistical method to relate the two for past forecast realizations. Then once the statistical relationship is developed, it is applied to future NWP model forecasts to infer the corresponding cloud amounts from the forecast states. Trajectory models use the cloud amounts or analyses to specify the cloud amounts at an initial time, and wind (and possible other parameter) predictions for the desired forecast times, and advect the moisture

field inferred from the initial cloud specification along the forecast trajectory for each initial location. Usually, a type of Lagrangian advection is used, and the moisture amounts carried forward in the process are converted back to cloud amounts to finalize the cloud amount forecasts. The AFGWC 5LAYER model operates in this manner. Finally, prognostic models attempt to predict cloud distribution explicitly by including predictive equations for clouds (usually, cloud water concentration) in the NWP model. The predicted cloud quantity may then directly or indirectly indicate the amount of cloud cover (cloud amount) at each forecast location. CldAmt94 included a discussion of past research in the development and testing of all three types of cloud cover prediction methods.

We have chosen at this time to develop the diagnostic method over the prognostic method for several reasons. First, beginning in the near future, AFGWC will begin receiving operational global NWP model forecasts from the Navy's Fleet Numerical Oceanographic Center (FNOC). The diagnostic method can be imposed on the forecasts received from FNOC, while the prognostic method would require major changes to the Navy's global NWP model. Secondly, it is clear to us from our review of prior diagnostic studies that consideration of potential predictors from NWP forecasts were limited to too few forecast output variables. This less computationally demanding alternative to prognostic cloud methods should be more fully explored to determine its limits of skill and usability. Finally, prognostic methods have not clearly been proven superior to diagnostic methods, primarily because the predictive skill limits of diagnostic methods have not been established. The U. S. Air Force is one of a very few organizations that require cloud amount predictions as an end product. Therefore, it is incumbent upon the Air Force to establish the baseline skill of diagnostic methods as a benchmark against which to evaluate the skill of prognostic schemes.

We have chosen to improve a statistical forecast method of relating observed fractional cloud cover (FCC) to observed or forecasted non-cloud meteorological fields as an alternative to trajectory cloud forecasting. There are several reasons why we think that this approach has greater potential than the trajectory approach

3

for cloud forecasting beyond 12 hours duration. First, most trajectory models incorporate very limited physical processes in the transport of moisture. In contrast, the state-of-the-art NWP models parameterize many physical processes related to moisture transport, even if they do not explicitly carry cloud water as a prognostic variable. If one installs the rather rigorous physical processes into a simple trajectory model, one would end up with a model not unlike the NWP models that use semi-Lagrangian advection rather than the Eulerian advection used in most NWP models. A recent study[5] designed to investigate the utility of semi-Lagrangian moisture transport showed that the Lagrangian method did not produce significantly better moisture forecast distributions. Secondly, random errors in wind forecasts grow quickly in NWP forecasts (even though average wind errors remain small), leading to large random errors in moisture displacement. Thirdly, the question of which moisture parameter to advect (degree of saturation, total water, water vapor concentration, etc.) is still open. In any case, the parameter must be truly conservative if conservation assumptions (apart from precipitation) are made in the trajectory method. Finally, the large-scale vertical motion is unlikely to be spatially representative of cloud motions, which is what is assumed when NWP model vertical motions are used in the vertical component of cloud trajectories.

We continue to feel that, in light of some of these uncertainties in the trajectory approach, the statistical forecast approach has more merit than we have been able to exploit to date. Thus, the scope of our current study involved pursuing the best possible representation of cloudiness from non-cloud meteorological fields. Ultimately, we expect that explicit prediction of cloudiness may surpass the best of diagnostic methods in forecast skill. It behooves the Air Force to establish a solid diagnostic method to act as a standard for such a comparison. Besides, the fact that the Navy will be supplying the global NWP products dictates that, until such time that prognostic approaches are clearly superior to diagnostic methods, there will be no justification for adding the additional burden to the Navy's weather prediction system.

Having chosen a statistical approach of relating cloud amount to meteorological fields, it is necessary to specify the three required components of this approach. This method requires a predictand, predictors, and the general and specific methods of relating the two. In the following sections, we describe the development and use of each of these components.

2. RTNEPH CLOUD AMOUNT AS THE PREDICTAND

The requirement for predictions of fractional cloud cover, or what we refer to as cloud amount, dictates that this quantity be used as the predictand in the development of the statistical method. This quantity is a measure of the fraction of grid area of the earth's surface obscured by cloud. The RTNEPH data base includes an estimate of the base and top altitudes of each cloud amount value reported, for up to four layers over a location.

Gridded analyses of cloud cover are favored over observational data in method development because of the differences between estimates of cloud cover from ground-based and space-based observing methods. The RTNEPH algorithm attempts to merge the differing sources into a single estimate of cloud amount. Also, the analyses provide estimates of cloud amount on regular spatial grids like the gridded weather parameters produced by the forecast models (to be used as predictors).

In the currently available cloud analysis archives, such as RTNEPH and the International Satellite Cloud Climatology Project[6,7] analyses, the uncertainty of the altitude assignments of the cloud bases and tops reinforces the essential two-dimensional nature of the cloud analyses. In CldAmt94, we dealt with this uncertainty by assigning the reported clouds in RTNEPH to prespecified altitude domains, which we called "decks" (high, middle, low). We continued to use that approach in this study, with the addition of a fourth deck, total cloud amount (amount of surface obscuration by any cloud). This quantity is provided directly in

the RTNEPH data set for each grid point, and is thought to be the most accurate quantity reported by RTNEPH.

As stated in CldAmt94, the RTNEPH's horizontal grid resolution of 25 nautical miles (about 48 km) true at 60 degrees latitude is more than adequate for the scales of motion resolved by today's operational global NWP models. A modest sample of hemispheric RTNEPH analyses provides a robust statistical sample of cloud scenes, even if it is limited to timely (no more than two hours old) observations. For practical reasons, we felt that it made sense to continue to use the RTNEPH cloud amount as the predictand (both in development and verification) because of our previous experience with the data, its availability in the operational environment, and our library of software to represent it on the grids compatible with NWP predicted fields. Thus, we used the January and July 1991 RTNEPH data sets previously obtained as a basis for developing our improved diagnosis schemes.

We used the horizontal and vertical transformation methods described in CldAmt94 to represent the cloud amounts on the so-called transform grid for compatibility with the NWP fields. This transform grid[8] is the European Center for Medium-Range Weather Forecasts (ECMWF) equal-area grid (approximately 125 km grid spacing) corresponding to a spectral triangular 106 wave forecast model in the horizontal, and the high, middle, and low cloud decks in the vertical. In the current study, we also placed the RTNEPH total cloud amount reports on the horizontal, equal-area grid.

In calculating the transformed RTNEPH cloud amount on the equal-area grid, we followed the practice of CldAmt94. Before the transformation, we assigned the high, middle, and low deck bases and tops by specifying the NWP model sigma levels (sigma = pressure/surface pressure) that define each deck interface. We then computed the monthly average altitude of each interface at each equal-area gridpoint from the ECMWF global meteorological analyses for the respective month. Then for each synoptic time in the month, we identified the nine (3 × 3) RTNEPH points centered on the point lying closest to the equal-area gridpoint. For each

RTNEPH point, we defined the identifying altitudes of the reported cloud layers as the reported base or top altitude for each layer depending on whether the report was based on a surface observation (base), a satellite observation (top), or radiosonde (both). For each of the nine points, we used the identifying base or top altitudes to assign them to the appropriate deck (H,M,L) for layered clouds in accordance with the deck base and top altitudes computed for each equal-area grid point. Any points with untimely data (more than two hours old) were not used. Then the simple average over the number of timely points (at least five, or the average was set to "missing") was computed for each deck and for total cloud. Any timely RTNEPH point that had no contribution to a particular deck or to total cloud was considered to contribute a 0 percent cloud amount to that deck or total cloud. The five-to-nine point average was assigned to the equal-area grid point as the transformed cloud amount.

One of the possible shortcomings of our previous representation of RTNEPH on the transform grid was the likely underrepresentation of high cloud and overspecification of middle cloud. In CldAmt94, we identified several possible reasons for this: (1) setting the base of the high cloud deck too high, (2) assuming high "thin" clouds to be zero cloud amount, and (3) possible underestimate of cloud altitude of high clouds in the RTNEPH. Correction of the latter problem was outside the scope of our efforts in the study. The possibility of setting the base of our high cloud deck too high could explain both too little high and too much middle cloud on the transform grid, while reason (2) would only explain the high cloud deficit.

We studied the possibility that we may have set the high cloud deck base too high in CldAmt94 by first computing the hemispheric average global spectral model (GSM) sigma level altitudes for both January and July 1991 from the ECMWF analyses. We found that the altitudes of the sigma level 10 (sigma = pressure/surface pressure = 0.45), our previous high deck base, were 6527 and 6772 m for January and July respectively. A histogram of RTNEPH identifying altitudes showed a large increase in frequency of occurrence at 6000 m in both months.

7

Above 6000 m, cloud altitudes are reported in 300 m increments. This suggests that clouds having base or top altitudes reported at 6000, 6300, or for July 6600m would have been assigned to the middle cloud deck in the transform grid.

The hemispheric average sigma level altitudes for sigma level 11 (sigma = 0.50) were computed to be 5781 and 5965 m for January and July 1991 respectively. Using sigma = 0.50 as our high deck base would insure that virtually all RTNEPH cloud amounts with identifying altitudes of 6000 m or higher would be placed in the high deck of the transform grid. Thus, for the present study we used sigma = 0.50 as the high deck base.

The corresponding altitudes of the middle cloud deck base (sigma = 0.80) for January and July 1991 are 2106 and 2184 m respectively. In looking at the reported cloud identifying altitude histogram, these sigma = 0.80 altitudes lie below altitudes where we found spikes in the histogram signifying default altitudes for middle cloud. Thus, we retained sigma = 0.80 (sigma level 17) as our definition of middle cloud base. We also kept our low cloud base set at sigma = 0.99 (sigma level 22).

We also conducted an investigation of the occurrence of clouds indicated as "thin" in the RTNEPH data set. We interrogated all timely data for 15 January and 15 July 1991, at both 0000 and 1200 UTC, looking for occurrences of thin cloud. We found an extremely low occurrence rate, averaging (for the July cases) about 100 counts of thin cloud out of over 31,000 counts of reported timely cloud amount. Upon looking at individual cases, we found occurrences of thin cloud corresponding to all three of our designated cloud decks. All occurrences found were identified as resulting from surface observations, visual satellite observations, or both. The actual reported thickness of the cloud layer (difference between top and base altitudes) was typical of clouds not reported as thin. From this, we concluded that the terminology "thin" may be a reference to optically thin. Because of the very low frequency of occurrence and the reported thickness of "thin" clouds, we chose to include all "thin" clouds in the transformed RTNEPH cloud amounts.

Thus, the complete set of changes to the RTNEPH transform algorithm includes: high cloud deck base changed from sigma = 0.45 to sigma = 0.50, "thin" clouds are now treated as any other reported cloud amount, and reported total cloud was transformed onto the RTNEPH transform grid (rather than resulting from a stacking of H, M, L clouds as in CldAmt94). In Figure 1, we display the monthly, zonal average cloud amount of the transformed RTNEPH using the revised algorithm (using 00, 06, 12, and 18 UTC times). These may be compared directly with the corresponding figures in CldAmt94 (Figures 6 and 7) resulting from the use of the original algorithm. The most obvious difference is the increase (decrease) of monthly, zonal average high (middle) cloud amount. The monthly, zonal mean high and middle cloud amounts now appear to be in better agreement with the earlier 3DNEPH layer cloud assessments (see CldAmt94, Figures 6 and 7). We attribute the decrease in middle cloud and increase in high cloud largely to lowering the high cloud deck base from sigma = 0.45 to sigma = 0.50. In Table 1, we show the revised cloud amount average and standard deviation monthly statistics for latitude zones and the hemisphere. A comparison with Table 3 of CldAmt94 highlights the changes in high and middle cloud amount statistics resulting from the transformation algorithm revision.

We also recomputed the frequency of occurrence of 5 percent cloud amount categories based on the transformed RTNEPH cloud amount using the revised transformation algorithm. Plots of revised frequency of occurrence for January and July 1991 are shown in Figure 2. In the frequency distribution plots, we find that in comparison to those of the original RTNEPH transformation (see CldAmt94, Figure 8) there is a much lower (15-20 percent) frequency of occurrence of 0 percent high cloud amount. This decrease is compensated by a slight increase in the frequency of occurrence in most cloudy categories in high cloud, especially below 30 percent cloud amount. Middle, low, and total deck frequencies of occurrence do not change as much from their previous values. Only the frequency of occurrence of 100 percent cloud amount in the middle deck shows an appreciable change, about 5-10 percent lower.
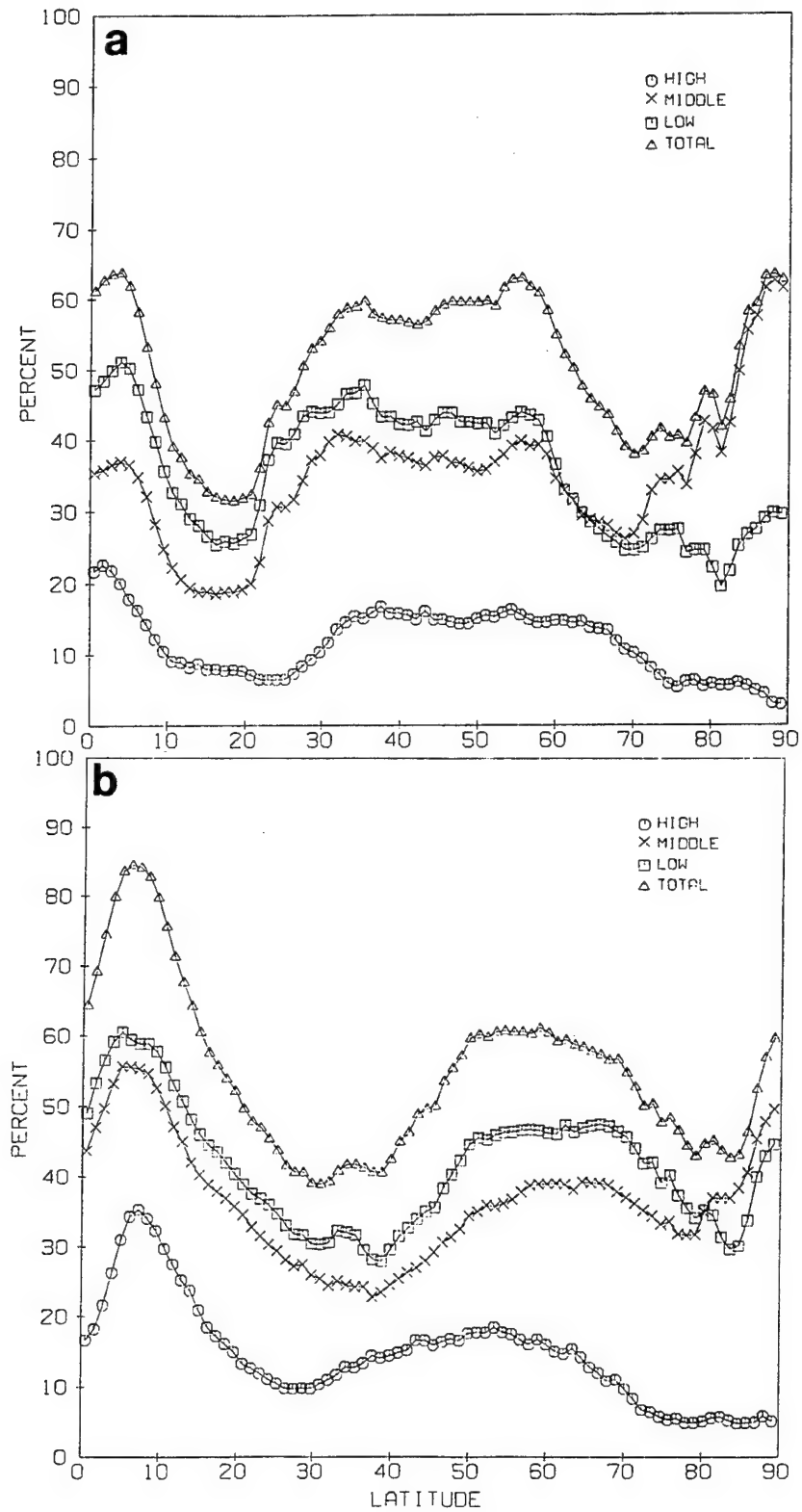
Figure 1. Zonal, Monthly Average Northern Hemisphere Cloud Amounts (Percent) for (a) January and (b) July 1991 Transformed RTNEPH.
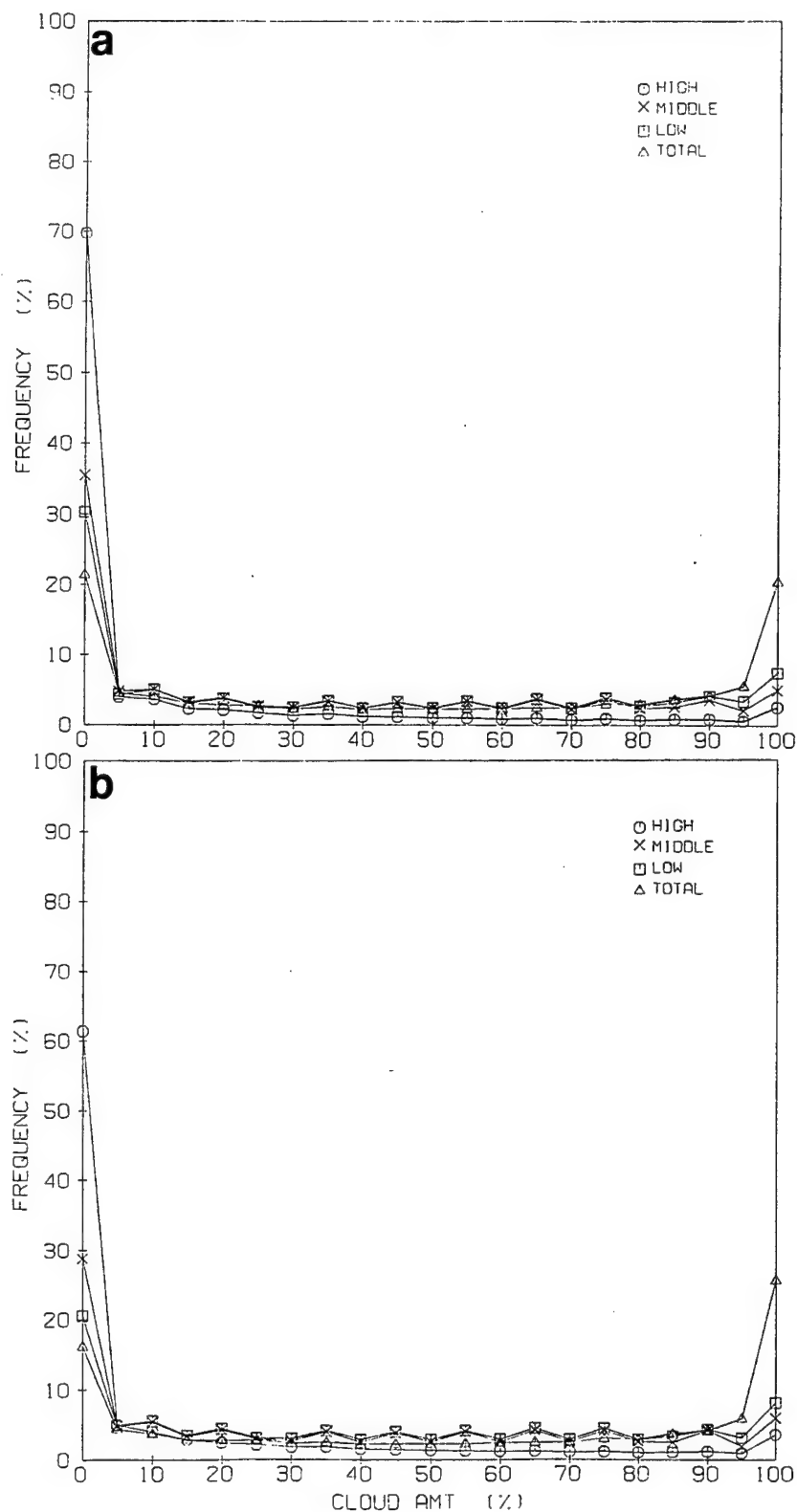
Figure 2.   Frequency of Occurrence (Percentage of All Transform Gridpoints for Four Times Per
Day Over the Month) of Cloud Amount in the Three Decks and Total Cloud for (a)
January and (b) July 1991.

Table 1. NH Transformed RTNEPH Cloud Amount Statistics

January 1991
Average Cloud Amount (Percent)

Cloud Deck	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
High	9.0	12.1	13.7	13.7	12.5
Middle	18.4	31.8	25.6	49.6	32.6
Low	23.0	43.5	27.5	51.0	37.9
Total	31.9	51.3	46.0	64.7	50.4

Cloud Amount Standard Deviation (Percent)

Cloud Deck	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
High	21.9	27.0	24.5	27.5	25.8
Middle	27.9	33.2	32.0	36.1	34.6
Low	30.4	35.3	34.0	37.1	36.4
Total	37.4	39.7	39.2	37.1	39.9

July 1991
Average Cloud Amount (Percent)

Cloud Deck	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
High	23.5	19.2	14.8	12.1	16.6
Middle	35.7	43.6	29.1	35.5	36.4
Low	36.8	50.3	37.2	41.0	42.5
Total	54.9	64.5	48.6	54.0	56.1

Cloud Amount Standard Deviation (Percent)

Cloud Deck	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
High	34.1	32.5	24.3	24.7	28.9
Middle	35.0	36.3	31.6	33.6	34.6
Low	34.4	35.6	34.3	34.0	35.1
Total	42.4	39.1	37.8	38.3	39.4

In the current study, we computed the standard deviation of the RTNEPH cloud amount about the five-to-nine point average value assigned to each equal-area grid point. We computed this quantity separately for each cloud deck and total cloud. We found that the standard deviation is higher on the edges of large cloud masses, and in the transition zones between cloudy and clear areas. The standard deviation values may give information about the degree of uncertainty of cloud diagnosis in an area of interest. They may also have possible application as a predictor in statistically diagnosing cloud amount. Time did not allow us to explore these possibilities in this project, but they may be worth considering in future refinements of the statistical diagnosis of cloudiness.

Another addition in this study was an attempt to use the reported altitudes of RTNEPH bases and tops more directly in relating vertical structure of observed cloud to vertical profiles of NWP variables. We concede that the uncertainty in altitudes discussed previously may very well prohibit a more precise vertical correspondence between reported cloud and a specific NWP model layer, for example. However, we felt that it may be of interest to see if horizontal correlations between cloud amount and cloud-related NWP variables might be improved through such a more precise vertical association. To that end, we stored the reported altitude of the highest top and lowest base of the five-to-nine points used in the average cloud amount assigned to the equal-area grid for each of the three cloud decks. We would like to use these altitudes in the future to see if there is any way to get higher spatial correlations between cloud amount and certain NWP variables using either NWP analyses or forecasts.

3. PL GSM FORECASTS AS THE PREDICTORS

As in the previous study, we chose to use forecast fields produced by the Phillips Laboratory Global Spectral Model[9] (PL GSM), a large-scale numerical weather prediction model. The PL-92 version was used in the previous study, which included the same physical parameterizations and numerical schemes as the

version described by Norquist and Chang.[10] Using this model afforded us the opportunity to seek improvements in the forecasts through model modifications. We could then rerun the cloud diagnosis using the improved forecasts to evaluate the sensitivity of the quality of cloud diagnosis to the quality of NWP model forecasts from which they are derived.

In both the previous and present studies, the PL-92 and modified PL-92 (designated PL-94) versions of the PL GSM were initialized using global meteorological analyses from the ECMWF. These analyses were acquired in spherical harmonic representation of a triangular spectral truncation of 106 waves (T106). Geopotential height, temperature, zonal and meridional wind components, and relative humidity available on 14 mandatory pressure levels were evaluated on a 320 longitude x 160 Gaussian latitude grid, interpolated to 22 model sigma layers, then transformed back to T106 spectral coefficients. In the previous study, these fields were then subjected to two iterations of a nonlinear normal mode initialization using the full model tendencies from PL-92. In the present study, we used the initialized T106 ECMWF analyses for both initial conditions and forecast verifications, so no further initialization step was necessary.

3.1 PL GSM Experiments to Formulate PL-94

The first step in the process of formulating the next improved version of the PL GSM was to identify single-forecast initial conditions that would result in forecast errors most representative of monthly error statistics. This was necessary because it was not feasible to execute each experimental version of the PL GSM from an ensemble of initial states representative of the month. To identify a representative initial state we initialized PL-92, our starting version of the model, from each of eight initial times in January and July 1991 spaced 3.5 days apart beginning with the first day of each month at 0000 UTC. We verified each 24-hour forecast against the initialized ECMWF analysis at the forecast valid time. We also verified the ensemble of eight forecasts against their respective valid time initialized ECMWF

analysis to compute ensemble errors. We then compared the individual forecast error statistics (mean error and root-mean-square error for u, v, T, Z, R H) for each forecast with the ensemble forecast error statistics for all mandatory pressure levels though 100 mb. The individual forecasts with the errors in closest agreement with the ensemble forecast errors were found to be those initialized on 11 January 1991 at 1200 UTC and 08 July 1991 at 0000 UTC. These initial times were chosen as the starting times for the experimental executions of the PL GSM.

The next step was to identify modifications to PL-92 that would have the potential to reduce the forecast errors. As with all NWP models, upgrades in formulation, surface condition specifications, and physical process parameterization are continually necessary to maintain a state-of-the-art model. Past and current studies of the PL GSM suggested a number of modifications that might improve the model.

In a simultaneous Phillips Laboratory-Atmospheric Sciences Division project, Yang et al.[11] developed and tested modifications to the PL GSM surface layer processes in a one-dimensional model. The current study offered an opportunity to test and evaluate the performance of these modifications in the full three-dimensional PL-92. Norquist and Chang[10] found that the model's systematic humidity error was particularly sensitive to surface evaporation rates realized in the model integrations. New land and ocean surface specifications were available, in both a new sea surface climatology and a modification to the plant canopy transpiration parameterization. Finally, in extending an investigation of the moisture bias in meteorological analyses to the ECMWF analyses used to initialize our model executions, we found that the ECMWF initialized analyses for January and July 1991 contained moist biases, particularly at low latitudes.

A series of four experimental versions of the PL GSM were executed from both the January and July selected initial times, and resulting forecasts were verified to compare their error statistics with those of experiment 0, named PL-92. In experiment 1 (named SSTMP), we changed the sea surface temperature specification from monthly average values from National Meteorological Center

climatologies on a 2.5 degree latitude-longitude grid to a more recent 1 degree latitude-longitude grid of monthly average values acquired from the National Center for Atmospheric Research. In experiment 2 (named OCNEV), we replaced the PL-92 formulation for surface roughness lengths over water surfaces with a scheme used in the current ECMWF global NWP model introduced by Miller et al.[12] This scheme effects the computed water surface fluxes of momentum, heat, and moisture into the atmosphere. In experiment 3 (called LNDMD), we modified a number of planetary boundary layer (PBL) parameterization formulations, including: land surface roughness lengths for heat and moisture, the computation of surface exchange coefficients for heat and moisture, the computation of the surface energy balance, snow cover effects on the computation of surface fluxes, and the computation of the soil heat flux. The details of these modifications are given by Yang et al.[11] In experiment 4 (called OCLND), we combined the modifications of experiments 2 and 3.

We used three references for the verification of the experimental forecasts. First, we computed zonal cross-sections of the mean error (ME) and root-mean-square error (RMSE) for both 24- and 96-hour forecasts when compared with the global ECMWF initialized analyses. Error categories were ME and RMSE for u, T, Z, and RH, for both 24- and 96-hour forecasts for both 11 Jan 91 1200 UTC and 08 Jul 91 0000 UTC (32 categories in all). Next, we computed error statistics for both 24- and 96-hour forecasts when compared with forecast valid time rawinsonde data in the latitude belt 30N-90N. Here, our error categories were ME and RMSE for Z, T, RH, and q (specific humidity), and vector wind RMSE. Counting 24- and 96-hour forecasts for each of two initial times, there were 36 error categories in all. Finally, we compared the global averaged and zonal averaged precipitation and evaporation rates from the various experimental forecasts with climatological values of precipitation of Jaeger[13] (8 error categories).

We compared the error values in the error categories mentioned above among the forecasts from experiments 0-4. We gave one point to the experimental version 0-4 that had the smallest error value in each error category. To do this, we had to

examine the errors at each mandatory pressure level for analysis and rawinsonde reference errors, and determine the experiment version that had the largest number of levels with the smallest error among the versions. When all error categories were evaluated and the points were tallied, we found the following results. Against the analysis reference, experiment 4 had more than twice as many points as any of the other experiment versions. In the rawinsonde reference verification, experiment 2 had a slightly larger point total than experiment 4, with the other versions lagging farther behind. And in the precipitation and evaporation evaluation, experiment 4 was the best in almost every error category. From these results, we concluded that the experiment 4 version of the model, OCLND, was to be the new reference against which another set of model modifications would be evaluated.

We executed a number of trial experiments consisting of various modifications of OCLND. We selected only three of them for a full-scale evaluation against OCLND. In experiment 5 (called NLSSP), we changed the plant canopy transpiration formulation. The quantity called canopy resistance was modified as follows: some vegetation parameter values were changed; parameter values were now set to represent a seasonal value based on ground surface temperature; canopy resistance is now computed rather than constant, and is dependent on weather and plant conditions, surface temperature, lowest model layer temperature, time of year, lowest model layer vapor pressure, albedo, and total solar radiation. In experiment 7 (called CRHUM), we modified the relative humidity of the initialized ECMWF analyses by replacing the zonal average of specific humidity of the analysis with the zonal average of specific humidity from a radiosonde climatology of Peixoto and Oort,[14] while holding the temperature analysis unchanged. This had the effect of reducing the water vapor content of the atmosphere, particularly at low levels in the tropics. In experiment 9 (called BLCLD), we replaced the low cloud amount specification formulation of Schattel[15] (a variant of the Slingo[16] cloud diagnostic scheme) with a formulation for boundary-layer cloud cover of Mahrt et al.[17]

We then compared the error values in the error categories used to compare experiments 0-4 to evaluate the best of the forecast versions among the verifications of experiments 4, 5, 7, and 9. We found that experiment 7, CRHUM, was by far the highest point-scorer against all three types of references. It was clear that the reduced atmospheric water vapor in the initial state led to smaller warm biases in the model forecasts, and this had beneficial impacts on the mass, motion, and moisture field forecasts. On this basis, we concluded that CRHUM, consisting of the modifications of sea and land surface roughness formulations, surface exchange coefficient formulations, surface energy balance, snow cover treatment, soil heat flux computations, and the water vapor content of the initial state, was the best of the experimental versions. The CRHUM version of PL GSM was thus designated PL-94, and served to produce the forecast fields from which the updated predictors were derived for the cloud amount diagnosis in this study.

We then executed the newly designated PL-94 version of the PL GSM from the same eight initial times in January and July 1991 as we had previously used to initialize PL-92. We verified both model versions' forecasts in ensemble against 30N-90N rawinsondes at 24-, 48-, 72-, and 96-hour forecast valid times. ME (or bias) and RMSE values were computed for Z, T, RH, and q, and RMSE for vector wind as in the experimental verifications. The results of the ensemble verifications of the PL-92 and PL-94 forecasts are shown in Figure 3-11 for the July forecasts only. We showed the July results because we verified in the Northern Hemisphere, and since most of the modifications involved physical processes that were expected to be more active in the summertime, we expected the impacts to be greater than in January. We found this to be true in the verification of virtually all variables. As can be seen in Figures 3-11, PL-94 forecast errors are smaller than PL-92 errors at most pressure levels and forecast times. This was true of the January forecasts also, but the difference between PL-94 and PL-92 errors were smaller in January than in July.

In comparing the forecast errors plotted in Figure 3-11, we see the most noticeable differences occurring in the ME and RMSE for Z and ME for T. In the
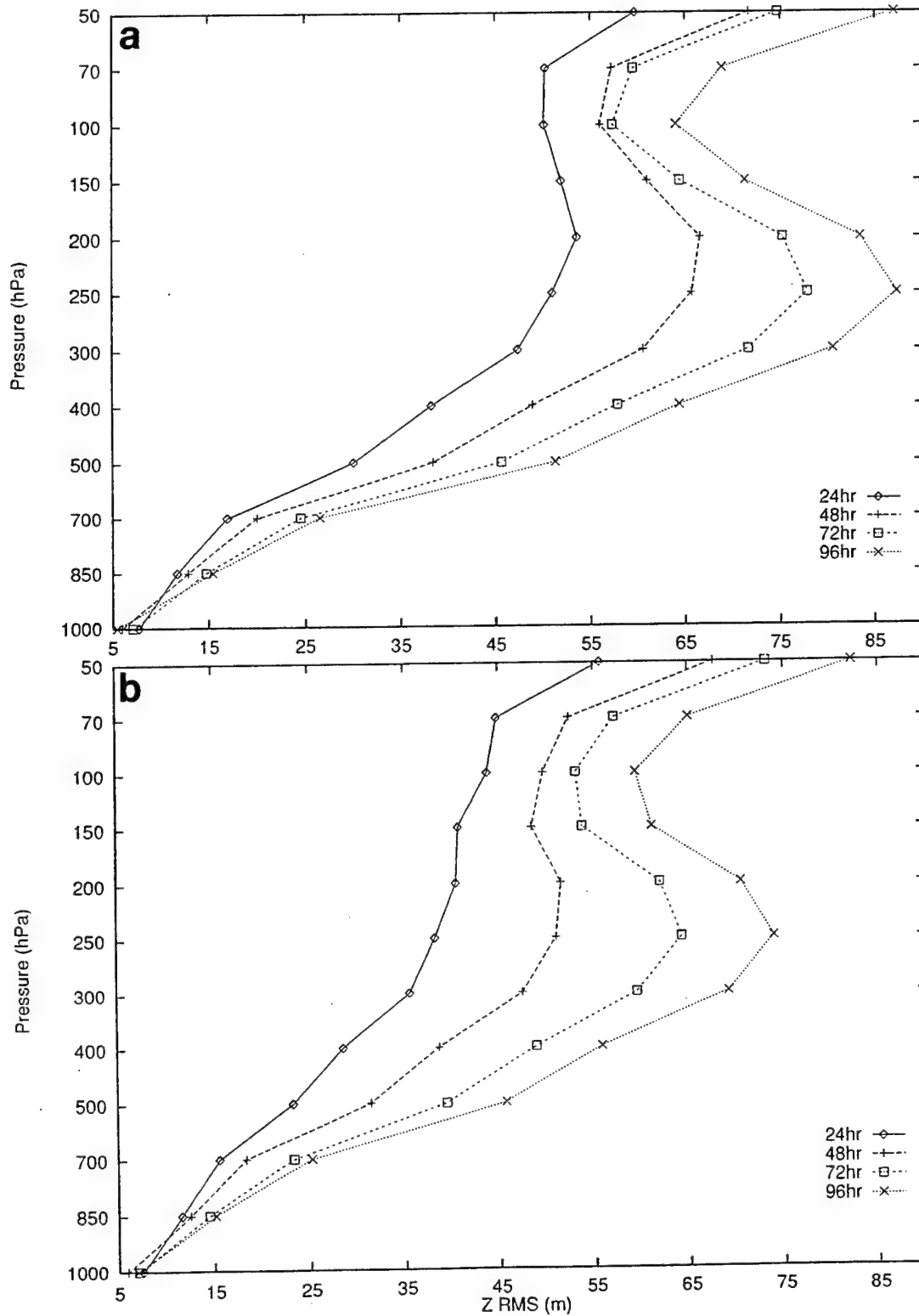
Figure 3. RMSE of Geopotential Height (a) PL-92 and (b) PL-94 Forecasts as Verified Against 30N-90N Rawinsonde Observations, Based on Eight Forecasts in July 1991.
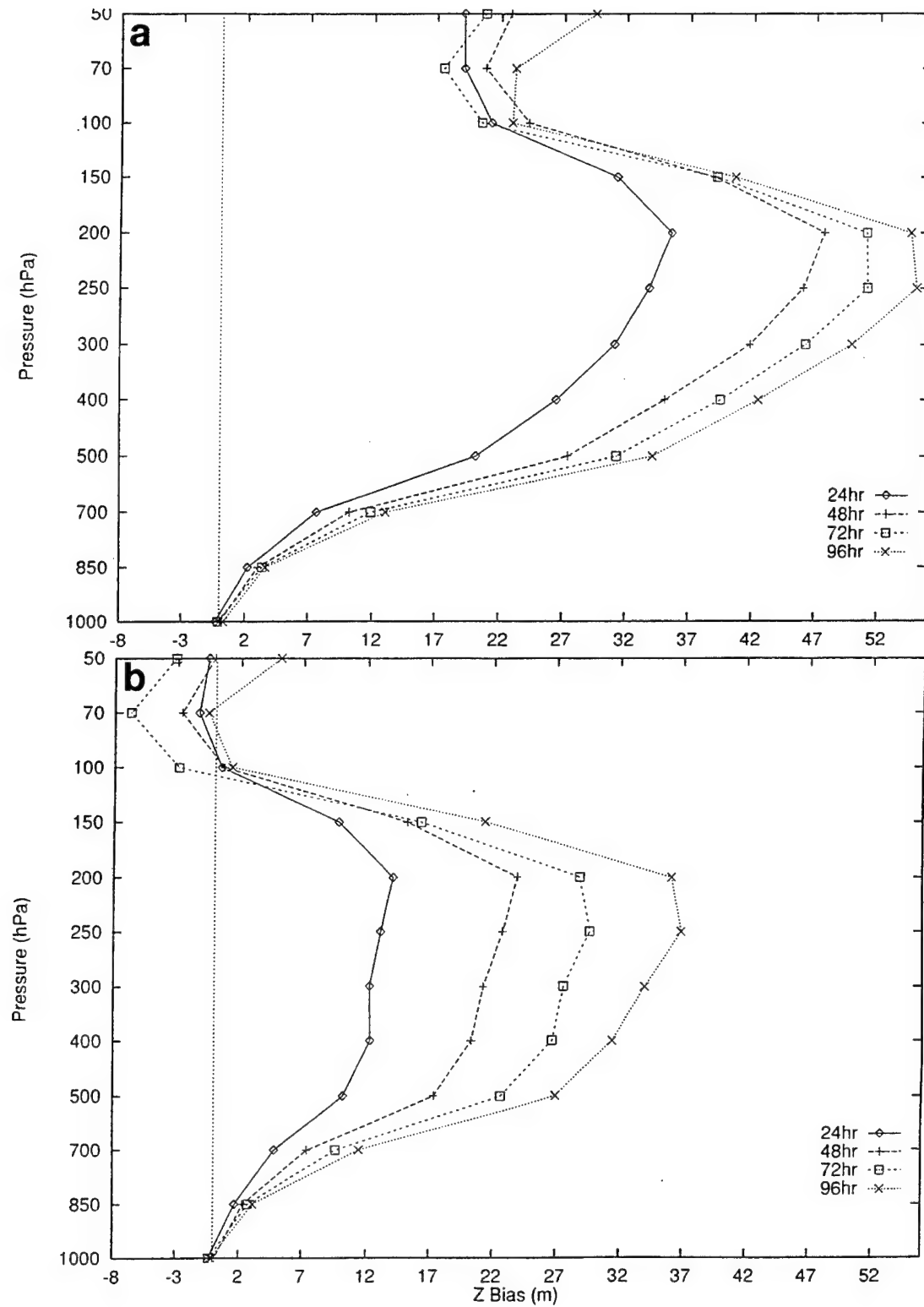
Figure 4.   Same as Figure 3 for Mean Error.

20
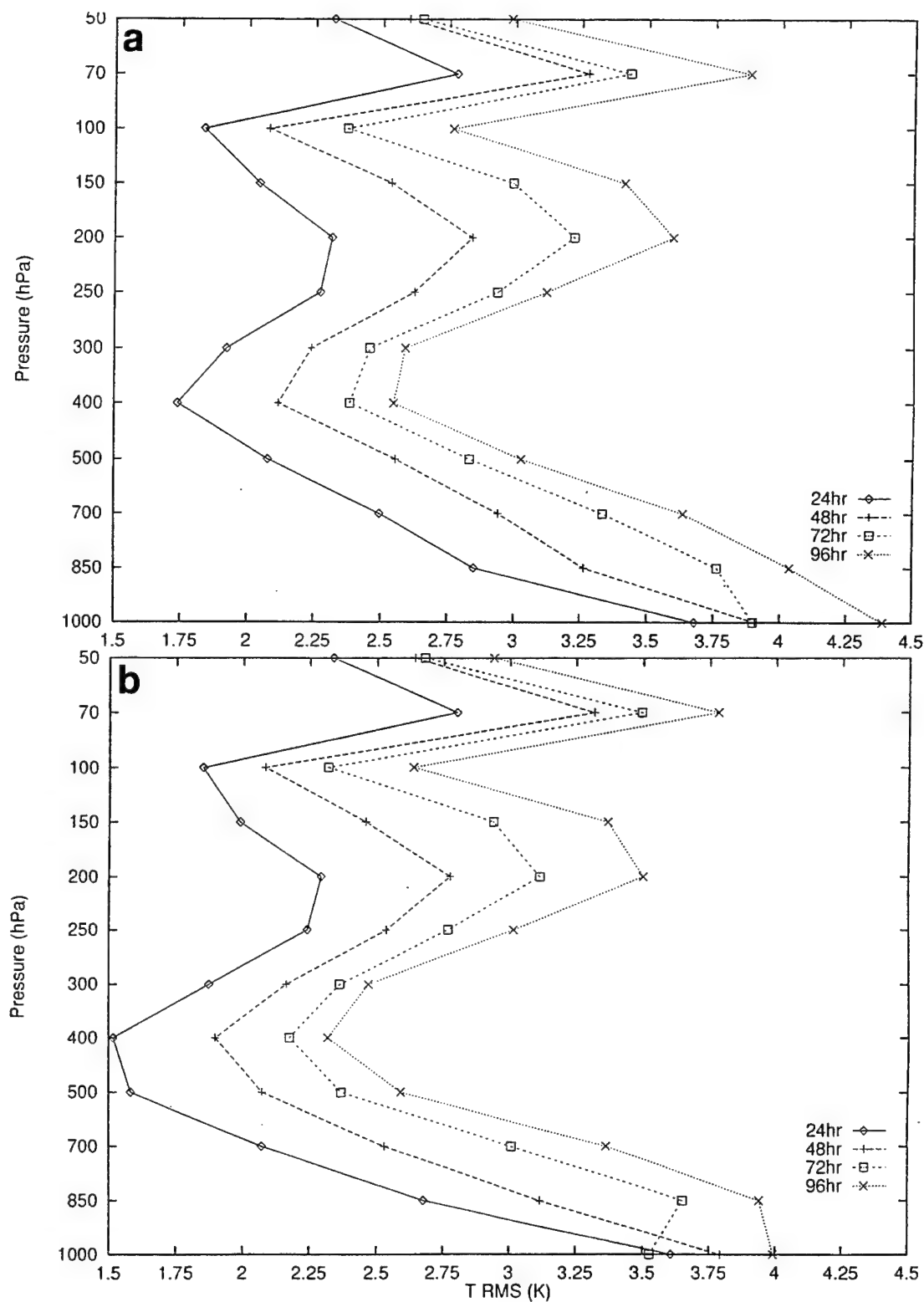
Figure 5. RMSE of Temperature (a) PL-92 and (b) PL-94 Forecasts as Verified Against 30N-90N Rawinsonde Observations, Based on Eight Forecasts in July 1991.
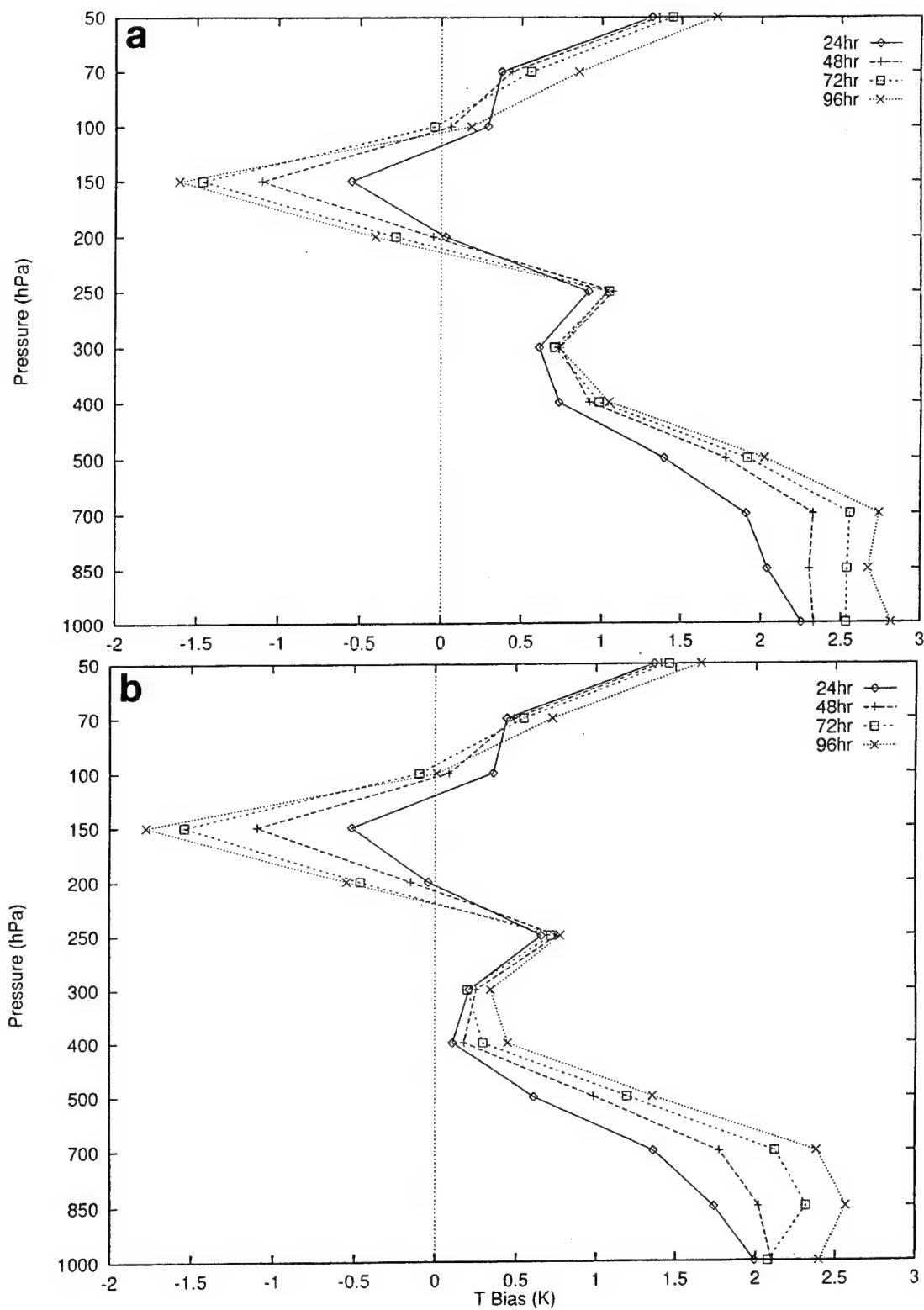
Figure 6.   Same as Figure 3 for Mean Error.
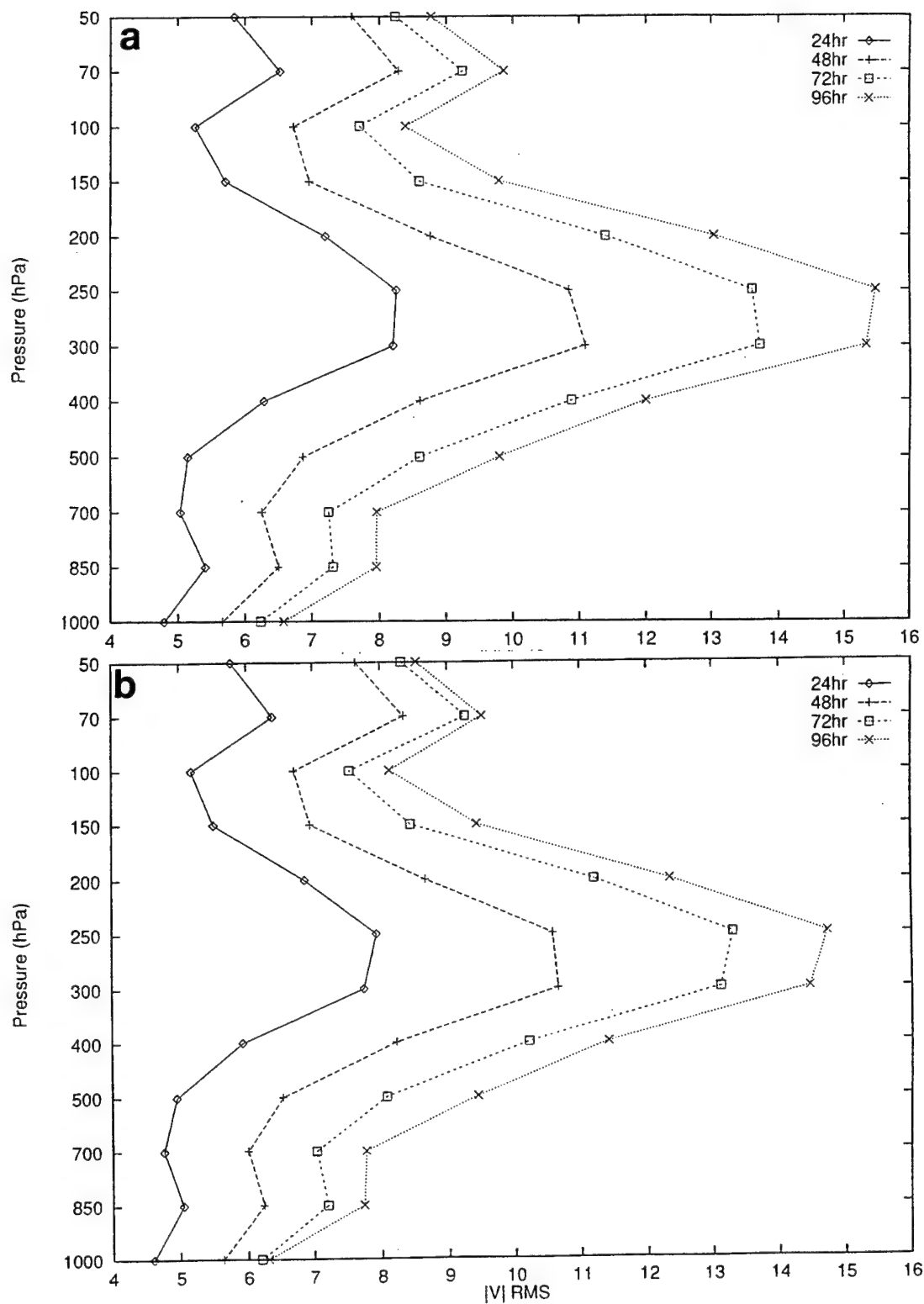
Figure 7. RMSE of Vector Wind (a) PL-92 and (b) PL-94 Forecasts as Verified Against 30N-90N Rawinsonde Observations, Based on Eight Forecasts in July 1991.
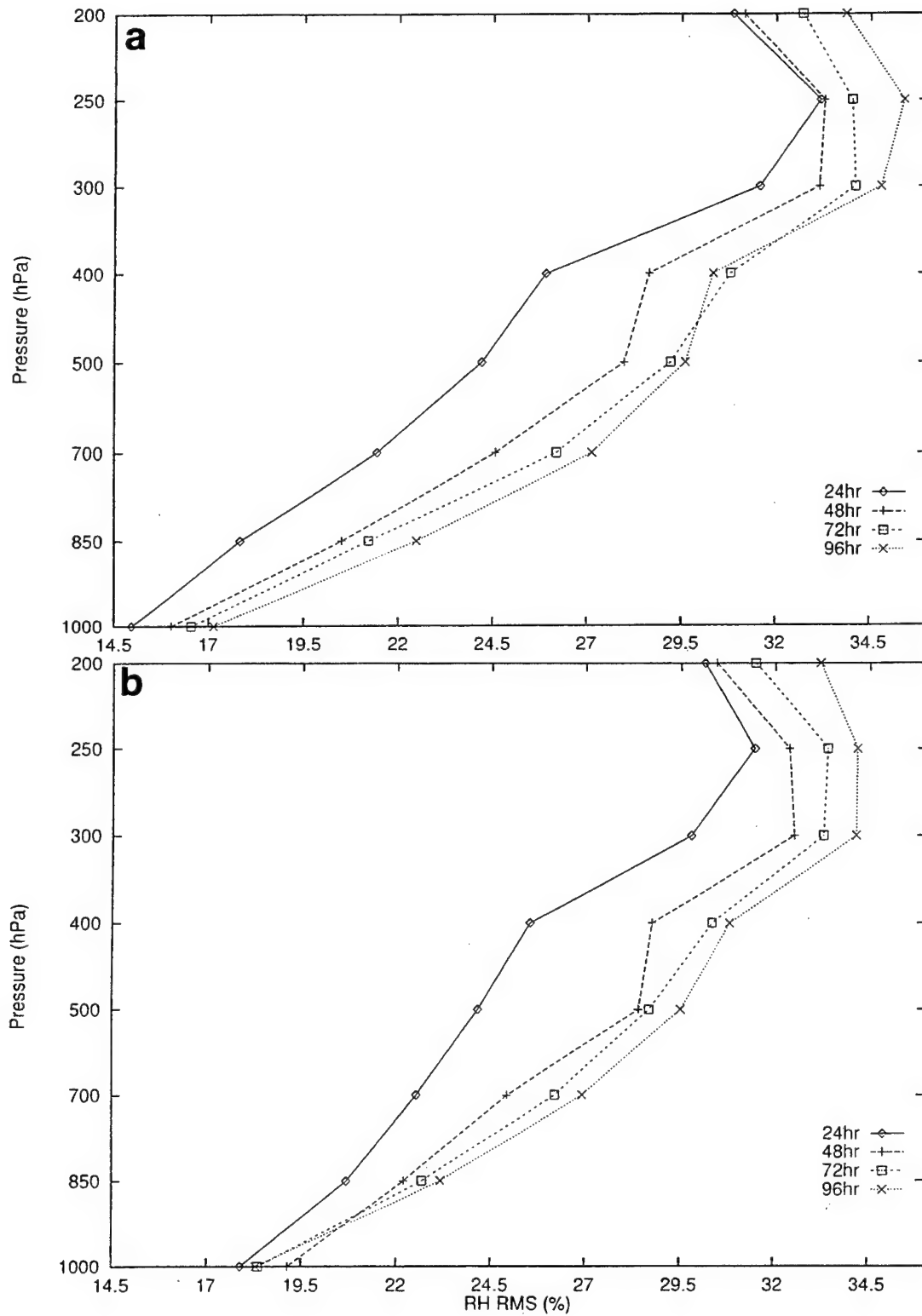
Figure 8.   RMSE of Relative Humidity (a) PL-92 and (b) PL-94 Forecasts as Verified Against 30N-
90N Rawinsonde Observations, Based on Eight Forecasts in July 1991.
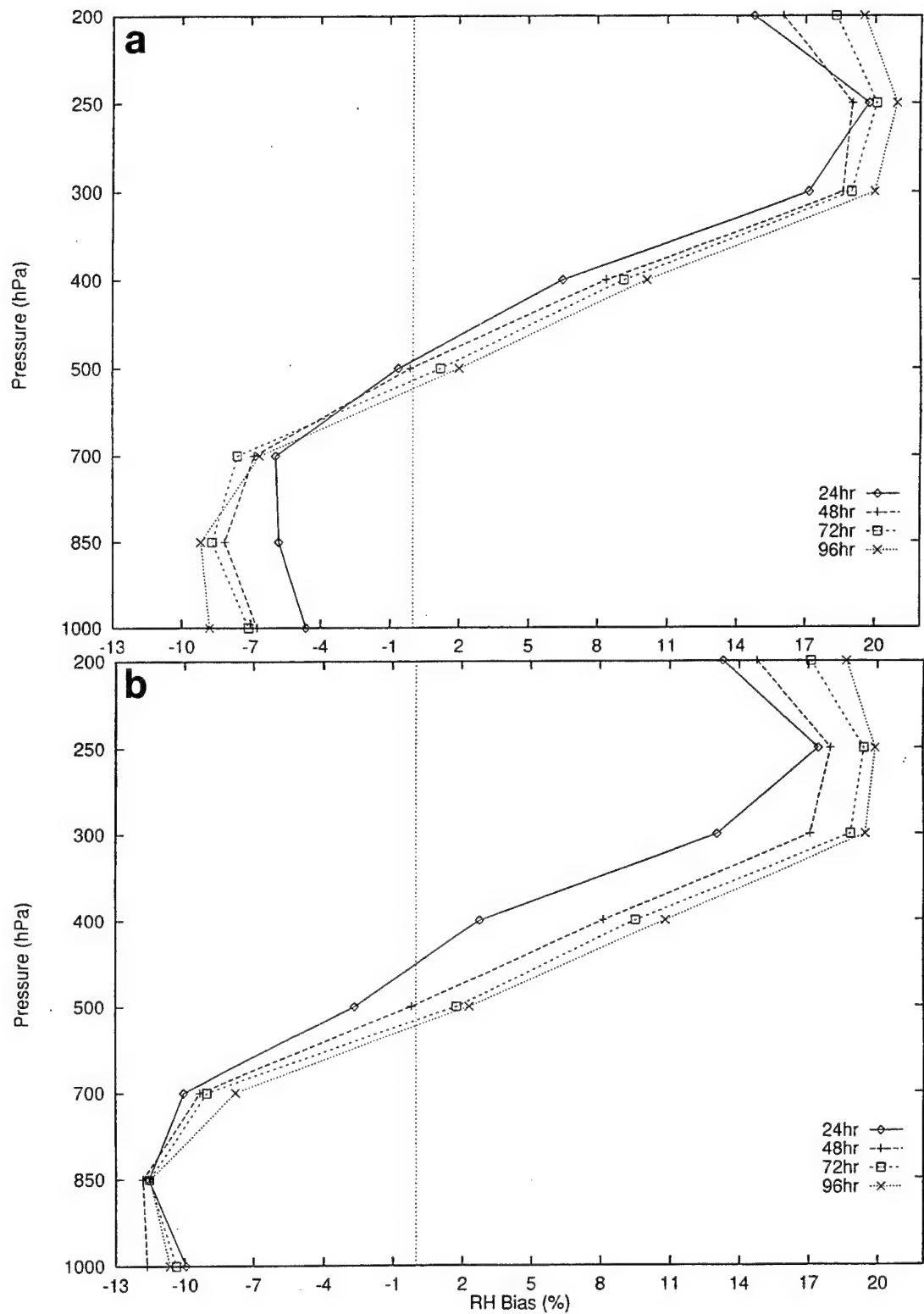
Figure 9.   Same as Figure 8 for Mean Error.
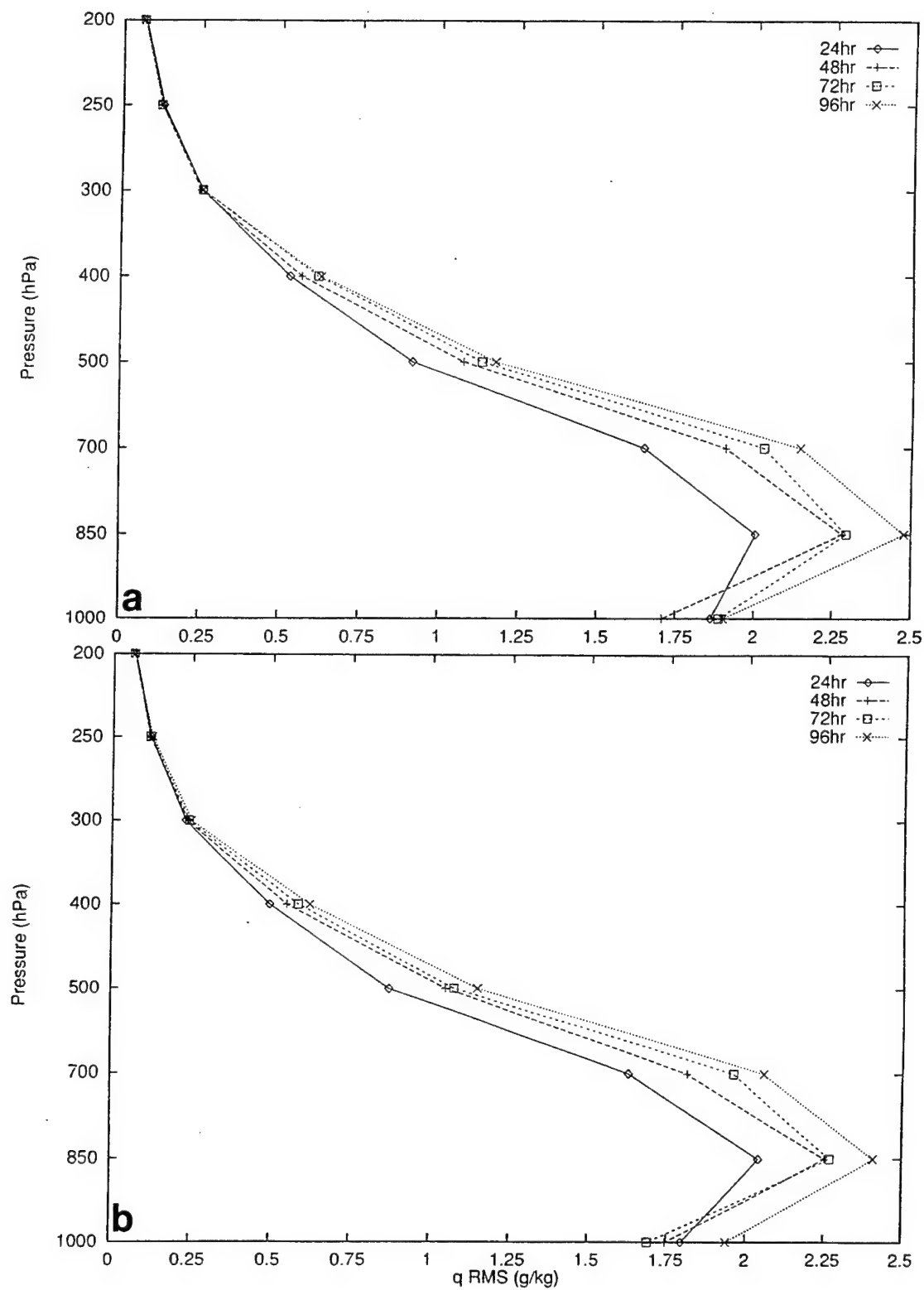
Figure 10. RMSE of Specific Humidity (a) PL-92 and (b) PL-94 Forecasts as Verified Against 30N-90N Rawinsonde Observations, Based on Eight Forecasts in July 1991.
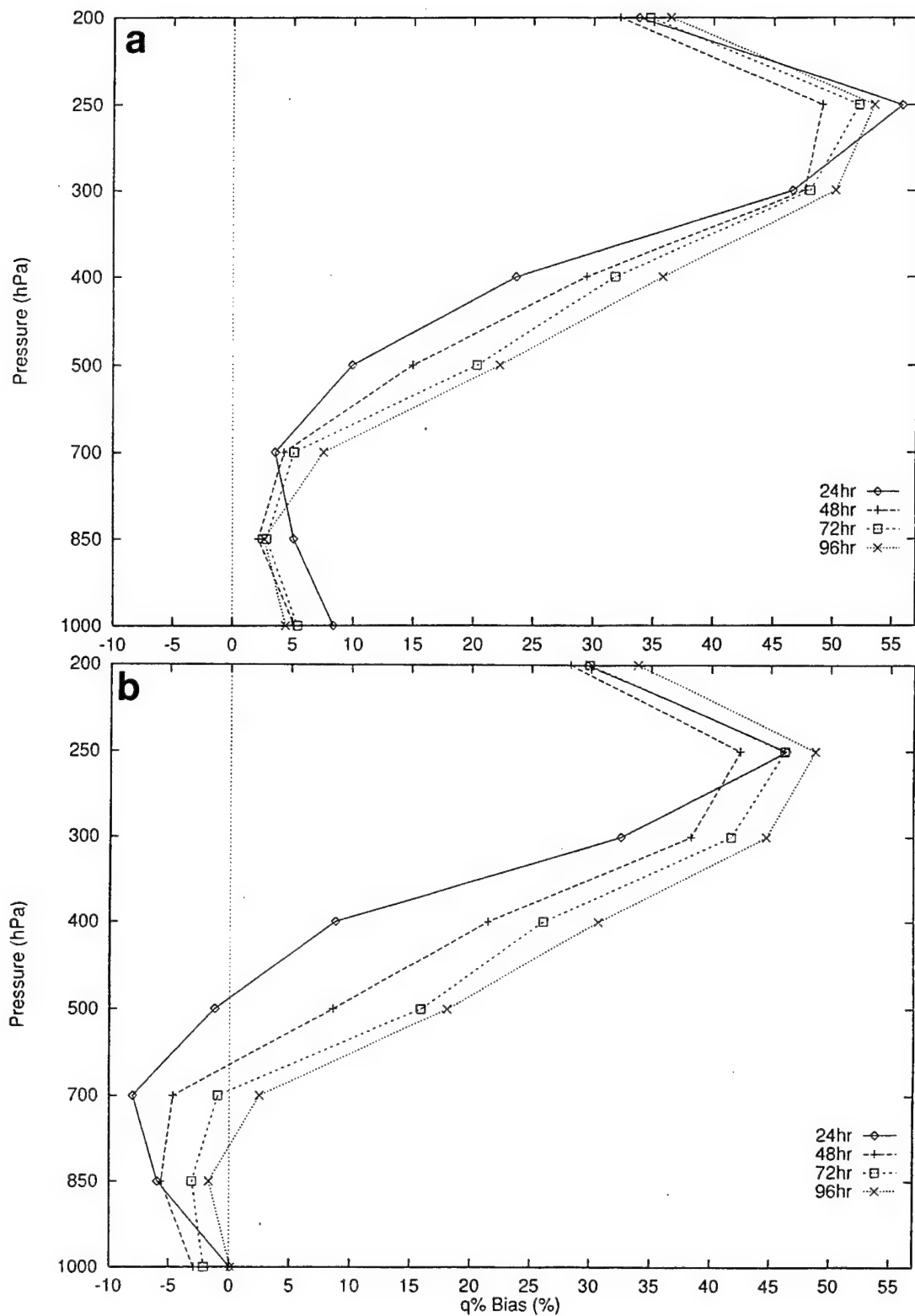
Figure 11. Same as Figure 10 for Percentage Mean Error (ME/averaged observed specific humidity).

latter, we find that the PL-94 warm biases at 250 hPa and below are consistently smaller than for the PL-92 forecasts, by as much as 0.5 K at some levels. This fact explains the reduced positive height biases in PL-94, which contributes significantly to the reduction of height RMSE. Vector wind error differences are modest, reaching a maximum of 0.5 ms^{-1} at the jet level. PL-94 RH RMSEs are actually slightly larger than their PL-92 counterparts at and below 700 hPa, and are just a few percent smaller above 700 hPa. This is likely due to the increased negative RH bias in PL-94 that is maintained through the length of the forecast at and below 700 hPa, and lower positive RH bias above 700 hPa in PL-94 forecasts. This same trend is even more pronounced in the q bias plots (Figure 11), where a small moist bias at and below 700 hPa in PL-92 is replaced by a dry bias in PL-94. Above that level, the PL-94 forecasts have a consistently smaller moist bias than do the PL-92 forecasts.

The trend toward a cooler, drier forecast model atmosphere in PL-94 can be ascribed in part to the magnitudes of the precipitation and evaporation rates produced by the respective models. In Figure 12 we show the ensemble, global averaged precipitation and evaporation rates for the respective model forecasts. The smaller precipitation rates in PL-94 are associated with the reduced warm temperature bias, likely a result of the reduced amount of latent heating in the formation of lesser amounts of rainfall in PL-94. The small evaporation rates in PL-94 are associated with the drier PL-94 forecast atmosphere, apparently through the reduced sources of water vapor to the atmosphere due to lesser evaporation rates. We found in our investigation of the effects of the forecast model modifications that experiment CRHUM had the biggest single impact in reducing precipitation rates, and experiment LNDMD had the biggest single impact in reducing evaporation rates. Apparently, these two effects were retained when the modifications were imposed together in PL-94.

As stated at the beginning of this section, the major motivation for seeking an improved forecast model in this project was the production of improved predictors for statistical cloud amount diagnosis. Using the very same predictor variables as

28
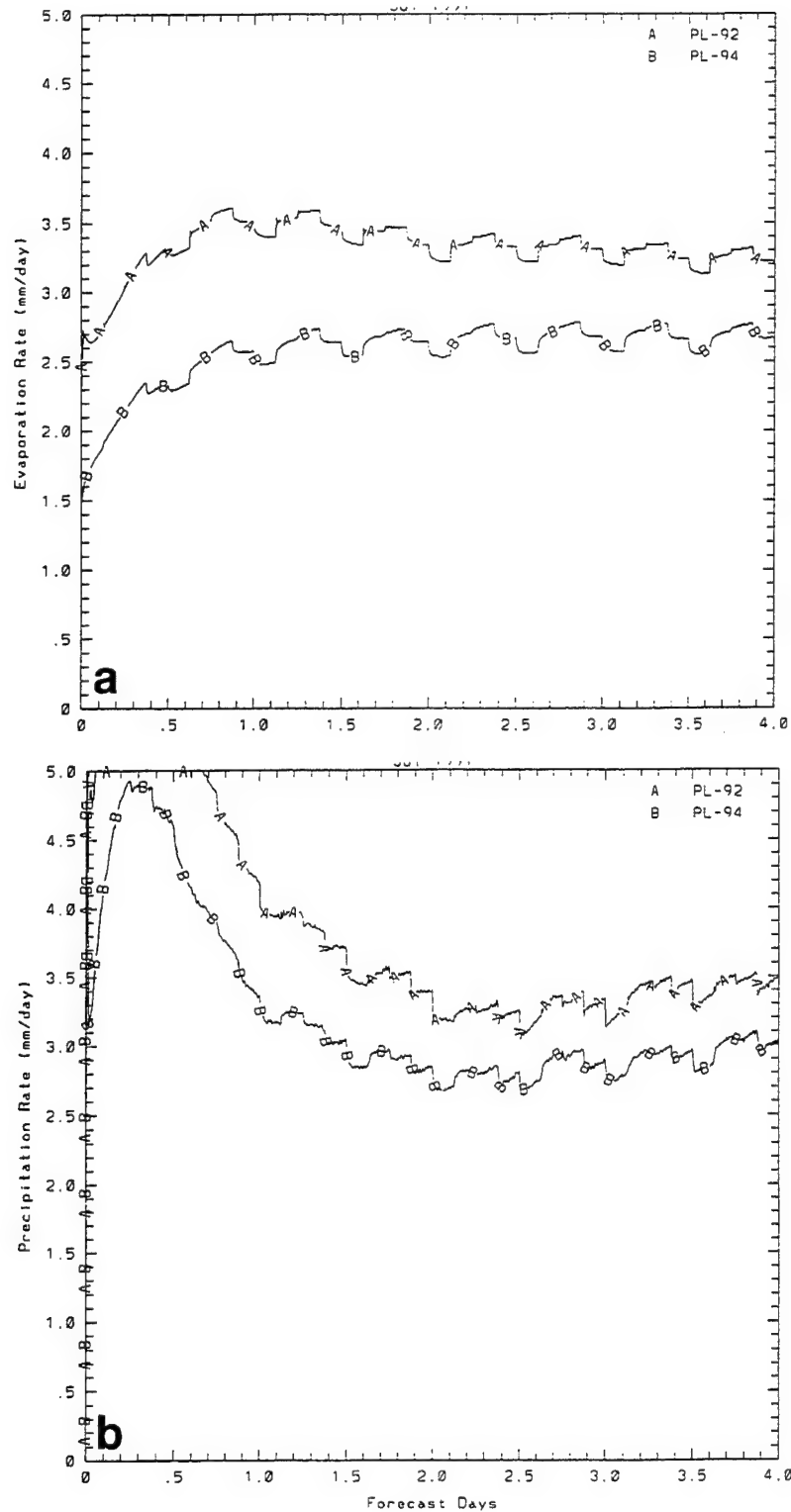
Figure 12. Global, Ensemble Average (a) Evaporation and (b) Precipitation Rates from Eight PL-92 and PL-94 Forecasts in July 1991.

29

in the previous study, but replacing their values with the PL-94 forecast values, affords the opportunity to investigate the sensitivity of cloud diagnosis from forecast states to NWP model forecast accuracy. We will quantify this sensitivity in a later section.

3.2 Computation of Cloud Diagnosis Predictors from PL-94 Forecasts

Our documentation of the previous study (CldAmt94) describes our approaches in the *a priori* selection of predictor variables. Briefly, the list of predictor variables we chose can be broken down into four categories of variables: dynamic, humidity, geographic, and turbulence. The dynamic variables would provide motion information to the cloud amount diagnosis. Such variables chosen include vorticity, divergence, three components of motion, and advection of vorticity, temperature, and water vapor. The humidity variables address both the spatial patterns of water vapor concentration (variables such as deck precipitable water) and the degree of saturation (variables such as RH and lifting condensation level). Geographic variables that are known to be associated with changes in cloudiness in a steady-state synoptic regime, such as hours of daylight left in the day and solar zenith angle, were included in the predictor list. Finally, many sub-grid scale clouds too small to be resolved by the global NWP model can only result from small-scale disturbances that we would call "turbulence". Large-scale indicators of areas of possible small-scale turbulence selected for predictors included static stability, wind shear, low level wind speed, and convective precipitation produced by the model.

To facilitate a comparison of cloud amount diagnosis skill between the two versions of the PL GSM, we used the same predictor list as in the previous study. This list is reproduced in Table 2 with the only changes being typographical corrections. Note that, unless otherwise indicated, the three-dimensional quantities are the pressure-weighted deck average at each equal-area gridpoint. In our case, this is the average of five model sigma layers in the low deck (0.8 < sigma < 0.99),

Table 2. List and Description of the "100-Predictors" Used in the
Selection of Multi-Linear-Regression Cloud Predictors.

No.	Name	Description
1	VORD6	Vorticity, predictand deck average, forecast t-6
2	DIVD6	Divergence, predictand deck average, forecast t-6
3	TMPD6	Temperature, predictand deck average, forecast t-6
4	PRWD6	Precipitable water, predictand deck average, forecast t-6
5	RHUD6	Relative humidity, predictand deck average, forecast t-6
6	OMGD6	Vertical velocity, predictand deck average, forecast t-6
7	STBD6	d(theta)/d(z), predictand deck average, forecast t-6
8	SPDD6	Wind speed, predictand deck average, forecast t-6
9	SHRD6	Wind shear, predictand deck average, forecast t-6
10	VADD6	Vorticity advection, predictand deck average, forecast t-6
11	TADD6	Temperature advection, predictand deck average, forecast t-6
12	QADD6	3-D humidity div., predictand deck average, forecast t-6
13	CPSD6	Condens. pres. deficit, predictand deck average, forecast t-6
14	MSTD6	d(theta-e)/d(z), predictand deck average, forecast t-6
15	UCMD6	West wind component, predictand deck average, forecast t-6
16	VCMD6	South wind component, predictand deck average, forecast t-6
17	RHXC6	Maximum RH within predictand deck, forecast t-6
18	RHAC6	RH at layer above maximum RH (see #17), forecast t-6
19	TMPC6	Temperature at maximum RH (see #17), forecast t-6
20	STBC6	d(theta)/d(z) at maximun RH (see#17), forecast t-6
21	SFCP6	Surface pressure (not sea-level), forecast t-6
22	RFST6	6-hr stratiform surface precipitation, forecast t-6
23	RFCV6	6-hr convective surface precipitation, forecast t-6
24	EVAP6	6-hr surface evaporation, forecast t-6
25	SPDB6	Surface-layer wind speed, forecast t-6
26	VORH2	Vorticity, high deck average, forecast t-0
27	VORM2	Vorticity, middle deck average, forecast t-0
28	VORL2	Vorticity, low deck average, forecast t-0
29	DIVH2	Divergence, high deck average, forecast t-0
30	DIVM2	Divergence, middle deck average, forecast t-0
31	DIVL2	Divergence, low deck average, forecast t-0
32	RHUH2	Relative humidity (RH), high deck average, forecast t-0
33	RHUM2	Relative humidity, middle deck average, forecast t-0
34	RHUL2	Relative humidity, low deck average, forecast t-0
35	OMGH2	Vertical velocity, high deck average, forecast t-0
36	OMGM2	Vertical velocity, middle deck average, forecast t-0
37	OMGL2	Vertical velocity, low deck average, forecast t-0
38	STBH2	d(theta)/d(z), high deck average, forecast t-0
39	STBM2	d(theta)/d(z), middle deck average, forecast t-0
40	STBL2	d(theta)/d(z), low deck average, forecast t-0
41	SPDH2	Wind speed, high deck average, forecast t-0
42	SPDM2	Wind speed, middle deck average, forecast t-0
43	SPDL2	Wind speed, low deck average, forecast t-0
44	SHRH2	Wind shear, high deck average, forecast t-0
45	SHRM2	Wind shear, middle deck average, forecast t-0
46	SHRL2	Wind shear, low deck average, forecast t-0
47	RHCH2	Maximum RH within high deck, forecast t-0
48	RHCM2	Maximum RH within middle deck, forecast t-0
49	RHCL2	Maximum RH within low deck, forecast t-0
50	TMPD2	Temperature, predictand deck average, forecast t-0

Table 2. (cont.) List and Description of the "100-Predictors" Used
in the Selection of Multi-Linear-Regression Cloud Predictors.

No.	Name	Description
51	PRWD2	Precipitable water, predictand deck average, forecast t-0
52	VADD2	Vorticity advection, predictand deck average, forecast t-0
53	TADD2	Temperature advection, predictand deck average, forecast t-0
54	QADD2	3-D humidity div., predictand deck average, forecast t-0
55	CPSD2	Condens. pres. deficit, predictand deck average, forecast t-0
56	MSTD2	d(theta-e)/d(z), predictand deck average, forecast t-0
57	UCMD2	West wind component, predictand deck average, forecast t-0
58	VCMD2	South wind component, predictand deck average, forecast t-0
59	RHAC2	RH for level above RH-max, predictand deck, forecast t-0
60	TMPC2	Temperature at maximum RH (see #17), forecast t-0
61	STBC2	d(theta)/d(z) at maximun RH (see#17), forecast) t-0
62	SFCP2	Surface pressure (not sea-level), forecast t-0
63	RFST2	6-hr stratiform surface precipitation, forecast t-0
64	RFCV2	6-hr convective surface precipitation, forecast t-0
65	EVAP2	6-hr surface evaporation, forecast t-0
66	SPDB2	Surface-layer wind speed, forecast t-0
67	RH2C2	Maximum-RH-squared within predictand deck, forecast t-0
68	RH4C2	Maximum-RH-fourth within predictand deck, forecast t-0
69	CCAC2	CCA cloud forecast, predictand deck, forecast t-0 (not used)
70	RHIC2	RH wrt ice at RH maximum, predictand deck, forecast t-0
71	LCDC2	Lifted-cond.-dist. at RH maximum, pred. deck, forecast t-0
72	LRIC2	Ln(Ri.-Number) at RH maximum. predictand deck, forecast t-0
73	GSLAT	Latitude (Gaussian grid, GS)
74	SGSLA	Sine of Latitude
75	CGSLA	Cosine of Latitude
76	SGSLO	Sine of Longitude
77	CGSLO	Cosine of Longitude
78	ZENA2	Solar zenith angle, forecast t-0
79	CZEN2	Cosine of solar zenith angle, forecast t-0
80	HRSS2	Hours of sunshine before forecast t-0
81	HRDK2	Hours of darkness before forecast t-0
82	SFCHT	Surface terrain height (9-pt ave., 1/8 mesh data)
83	SDVHT	Standard deviation of surface terrain height
84	PCH2O	Percent of surface that is water (from 1/64 mesh data)
85	DZ/DX	Eastward gradient of terrain height
86	DZ/DY	Northward gradient of terrain height
87	LRN92	3x3x3 (ijk) minimum of Ln(Ri.-Number), forecast t-0
88	STN92	3x3x3 minimum of d(theta)/d(z), deck average, forecast t-0
89	SHX92	3x3x3 maximum of vertical shear, deck average, forecast t-0
90	SPX92	3x3x3 maximum of wind speed, deck average, forecast t-0
91	RCX92	3x3 maximum of 6-hr convective rainfall, forecast t-0
92	SBX92	3x3 maximum of surface layer wind speed, forecast t-0
93	SVX92	3x3 maximum of surface-speed-times-terrain-var., frcst t-0
94	WBLL2	Surface wind times terrain gradient, forecast t-0
95	ABTV2	Minimum of terrain ht. or wind/stability height, frcst t-0
96	RH2D2	RH-squared, predictand deck average, forecast t-0
97	RH4D2	RH-fourth, predictand deck average, forecast t-0
98	RH2D6	RH-squared, predictand deck average, forecast t-6
99	RH4D6	RH-fourth, predictand deck average, forecast t-6
100	CLDOB	Predictand, observed RTNEPH deck cloud cover, forecast t-0

six sigma layers in the middle deck ($0.50 <$ sigma < 0.80), and 4-6 sigma layers in the high deck [sigma (lat) $<$ sigma < 0.50, where sigma (lat) $= 0.20$ for lat $= 0$-20N, sigma (lat) $= 0.25$ for lat $= 20$-65N, and sigma (lat) $= 0.30$ for lat $= 65$-90N].

The twice-daily PL-94 forecasts for days 8-24 of January and July 1991 were transformed to the equal-area grid in accordance with the three cloud decks. The only difference from the previous study in the construction of the predictor values on the equal-area transform grid was the earlier choice of sigma $= 0.45$ as the high deck base. Thus, the PL-92 predictors from the previous study included the sigma layer $0.45 <$ sigma < 0.50 in the middle deck. Even though the revised transformed RTNEPH cloud amounts now are based on a high deck base of sigma $= 0.50$, we didn't feel that the inclusion of this sigma layer in the middle deck average predictor computation (and the lack of it in the high deck computation) would require a recomputation of the transform grid PL-92 predictor values. In both the PL-92 and PL-94 predictor value data sets, the predictors were available on the equal-area transform grid, Northern Hemisphere only, at 6-h forecast intervals out to 48 hours of forecast time. As shown in Table 2, a number of the predictor variables are made available to the cloud diagnosis scheme at both the forecast diagnosis time and at a forecast time six hours earlier. This facilitates the inclusion of trend information into the cloud diagnosis scheme.

Finally, a new set of predictor values for total cloud was computed in the present study. In the previous study, the deck cloud amounts were diagnosed, and these were stacked to yield an estimate of total cloud. In this study, we used the RTNEPH-reported total cloud as a separate predictand, and prepared a separate set of predictors for its diagnosis. To do this, we used the pressure-weighted average of the three deck averages for all predictors that were deck-specific in each of the individual deck predictor lists. All other non-deck specific variables, and variables which had information for all three decks, were common to the high, middle, low, and total predictor lists. In the balance of this report, total cloud will be considered the fourth cloud deck.

33

4. CLOUD AMOUNT DIAGNOSIS USING MULTIPLE LINEAR REGRESSION

The multiple linear regression (MLR) scheme developed and tested in CldAmt94 was unchanged in this study. This included the use of the IMSL STAT/LIBRARY software routine called DRSTEP with a forward stepwise selection method, a limit of the 20 top predictors selected, and a factor of 2 applied to the computed regression slope. The latter represents a compromise between minimized mean-square error (using the computed slope) and preservation of the cloud amount frequency of occurrence distribution (which is better preserved with a steeper slope).

As in the previous study, we had only the timely 3×3 average RTNEPH cloud amounts at each equal area gridpoint. This represented some 50-60 percent of the total number of equal area gridpoints (just under 20,000 in the hemisphere). In computing the cloud amount-to-predictor statistical relationship, we randomly selected only one out of every four available equal area points in the hemisphere at each synoptic time. This reduction of the sample size was an attempt to insure a better statistical independence in the sample. As a result, there were generally 2000-2500 equal area grid points of cloud amount and predictor values making up the sample at each synoptic time. As in the previous study, we used a 10-day development period for the computation of the statistical relationship between cloud amount and the selected predictors. Since there were 20 such times in each 10-day development period, we ended up with a total sample size for the MLR of about 45,000 points for each cloud deck.

A single cloud amount-to-predictor set relationship was developed in each deck for the Northern Hemisphere using the 10 days of twice-daily n-hour forecasts and the corresponding transformed RTNEPH cloud amounts at the forecast valid times, for each of n=12, 24, 36, and 48 hours. This was done separately for each of the seven 10-day periods beginning with day 8 (that is, days 8-17) and ending with day 14 (that is, days 14-23) of both January and July 1991. Then for each of these

seven 10-day development periods, we applied the resulting statistical relationships to the predictors from forecasts initialized on the day following each 10-day period (in our case, the 18th through the 24th). The resulting diagnosed cloud amounts are thus valid for the n-hour forecast (n = 12, 24, 36, or 48) initialized on 0000 and 1200 UTC on each of the dates 18-24 January and July 1991. These two 7-day periods will hereafter be referred to as our verification periods.

The forward stepwise selection method prioritized the leading predictors in each of the seven 10-day development periods. The leading predictors are selected on the basis of highest correlation (for the first selectee) and multiple correlation with cloud amount. To draw attention to the most highly contributing predictors, we devised a simple scoring scheme in which a variable identified in the top 10 predictors in a given development period was assigned two points, and variables in the second 10 predictors are assigned one point. We added the point assignments over the seven development periods and designated two predictor categories: strong predictors (total points $\geq$ 10) and useful predictors (10 > total points $\geq$ 5).

In Table 3, we list the strong and useful predictors for each deck of diagnosis resulting from the use of 12-hour PL-94 forecasts for the seven 10-day development periods in both months. For comparison, the corresponding table for 12-hour forecasts is given as Table 7 in CldAmt94. Many of the top contributing predictors are the same in PL-92 and PL-94, but there are some notable exceptions. The dynamic variables: vorticity, vertical velocity, and moist convergence disappear from the PL-94 table. Relative humidity and temperature at the model layer of maximum relative humidity has now appeared in the PL-94 table. Relative humidity and dry static stability remain strong as predictors, while precipitable water, temperature, percent surface water are among variables that grow in importance, and $3 \times 3 \times 3$ minimum stability and $3 \times 3 \times 3$ maximum wind speed lessen in their contribution. Over all four decks, the strongest predictors of cloud amount from 12-hour PL-94 forecasts are: January--relative humidity, relative

Table 3.   Strong (x) and Useful (+) Predictors in 12-Hour PL-94 Forecasts As Determined by MLR Forward Stepwise Selection Method

JANUARY

Dynamic	L	M	H	T
temperature	+		+	x
zonal wind			+	
meridional wind	+			x
surface pressure	x			+
temp @ max RH lvl			x	

Humidity	L	M	H	T
relative humidity	x#	x~	x	x~
(rel humid)^2	x		x	x
(rel humid)^4	x	+	+	x
precip water	+		x	
ice rel humid	+		x	
cond pres spread			+	
lifting cond level	+			
(rel humid max)^2	+			
rel humid max				+

Geographic	L	M	H	T
sin (longitude)	+	x~	x	x~
cos (longitude)			x	x
hrs sun before t	+		+	x
hrs dark before t		+	x	x
% surface water	x	x	+	
cos (latitude)				
Z* stnd dev	+			+

Turbulence	L	M	H	T
dry static stab	x*	x*#		x~
moist static stab	+	+		+
wind speed		+*	x*@	+
wind shear			+	+
ln (Ri) @max RH lvl	+	x	x	+
3X3X3 min stab	+	x		
3X3X3 max speed	x	+		+
3X3 sfc max speed	+	x		
evaporation rate		x		
conv prcp rate			+	
Z* min/wind-stab Z		+		+
sfc layer wind spd				+

JULY

Dynamic	L	M	H	T
temperature	x		+	x~
meridional wind	+		x	+
stratiform prcp	x		x	
temp @ max RH lvl	+			

Humidity	L	M	H	T
relative humidity	x@	x	x@	x~
(rel humid)^2		+	+	+
(rel humid)^4	x	x	x	+
precip water	x	+		x
cond pres spread	x			
rel humid max	x#	x@#	+	x#

Geographic	L	M	H	T
sin (latitude)	x@	x	x@	x~
latitude	+	+	+	+
sin (longitude)	+	x	x	+
cos (longitude)		+	x	+
cos (zenith angle)			x	x
hrs sun before t	x		+	x#
hrs dark before t	+	x	x	+
% surface water	x	+	+	
surface height	+		+	
zenith angle	x			x

Turbulence	L	M	H	T
dry static stab	+~	x		x@#
wind speed	+	+*	+	+#
wind shear	x	+	x	
3X3X3 max speed		x	+	x
3X3 sfc max speed	x	x	+	x
evaporation rate	x	+	x	
sfc layer wind				

Predictors are from the deck of diagnosis except as follows:
* low deck; @ middle deck; # high deck; ~ all 3 decks

Note: No distinction is made here between t-6 and t-0 values

humidity squared, dry static stability, relative humidity to the fourth power, percent surface water, and 3×3 surface maximum wind speed; July--relative humidity, relative humidity maximum, evaporation rate, precipitable water, hours of darkness before forecast time, and dry static stability.

5. SENSITIVITY OF MLR CLOUD DIAGNOSIS TO DIFFERENCES IN PREDICTOR VALUES

We diagnosed cloud amount from the 12- and 48-hour forecast predictors of both PL-92 and PL-94 for the 7-day verification periods in both January and July 1991. We then verified the resulting diagnosed clouds against the forecast valid time transformed RTNEPH cloud amounts at all equal-area points having sufficient (5-9) timely 3×3 average RTNEPH cloud amount reports. We computed such skill scores as bias, RMSE, mean absolute error, and 20/20 score over the 14 synoptic times in the 7-day verification periods. The results of these verifications are shown in Table 4.

In the January 12-hour forecasts, the diagnosis skill is slightly but consistently better in PL-94 forecasts. However, the consistency of the trend from deck to deck and score to score does not occur in the January 48-hour forecasts. The July 12-hour forecast diagnoses show even less of a difference in skill between PL-94 and PL-92 forecasts, and less consistency than the January 12-hour forecasts. The July 48-hour forecast skill score differences consistently favor PL-94 forecasts but by a similarly small amount.

To evaluate the impact that a change in NWP model forecast skill has on the skill of cloud amounts diagnosed from those forecasts, we present the percent change in RMSE between the 48-hour forecasts of PL-94 and PL-92 in Table 5. We chose 48-hour forecasts to evaluate this sensitivity to allow sufficient growth in the difference between the forecasts of the two model versions. We computed the temperature, wind, and relative humidity RMSE of both model versions separately

Table 4. Verification Scores for MLR Diagnosis of Cloud Amount, 12- and 48-Hour Forecasts from PL-92, PL-94 for the Verification Periods 18-24 January and July 1991, Northern Hemisphere.

Deck	January					July			
	12-Hour		48-Hour			12 Hour		48 Hour	
	PL-92	PL-94	PL-92	PL-94		PL-92	PL-94	PL-92	PL-94
Bias (% Cloud Amount)									
H	3.37	3.24	1.85	2.11		5.10	5.31	3.88	4.66
M	0.45	1.00	0.78	0.82		2.99	3.25	2.33	2.79
L	0.78	1.34	1.39	1.11		2.45	2.32	1.91	2.16
T	-4.41	-3.41	-2.71	-3.00		-1.40	-1.30	-1.06	-0.89
RMSE (% Cloud Amount)									
H	25.17	24.33	25.53	26.84		28.76	28.29	29.86	29.45
M	33.18	32.56	33.91	33.38		34.14	34.20	34.65	34.55
L	33.50	32.94	34.17	34.17		34.56	34.28	34.99	34.64
T	34.89	33.96	36.47	35.95		33.48	33.40	35.65	34.46
MAE (% Cloud Amount)									
H	15.12	14.58	16.20	16.69		18.43	17.96	19.86	19.24
M	24.54	24.27	25.55	25.17		26.34	26.20	27.47	27.07
L	25.13	24.85	26.11	26.19		27.26	26.91	27.85	27.35
T	25.29	24.70	27.48	26.80		23.58	23.17	26.14	24.71
20/20 Score									
H	0.74	0.75	0.73	0.72		0.68	0.69	0.65	0.67
M	0.56	0.56	0.53	0.54		0.51	0.51	0.47	0.49
L	0.53	0.53	0.51	0.50		0.48	0.49	0.47	0.48
T	0.55	0.56	0.50	0.52		0.58	0.59	0.53	0.56

Table 5. Percent change in RMSE from PL-92 to PL-94 for Ensemble January and July 48-Hour Forecasts of T, V, and RH, and for the 7-day Verification Period for Cloud Amount.

Deck	January 1991				July 1991			
	ΔRMSE(T)	ΔRMSE(V)	ΔRMSE(RH)	ΔRMSE(CA)	ΔRMSE(T)	ΔRMSE(V)	ΔRMSE(RH)	ΔRMSE(CA)
H	-2.9	-1.0	-2.1	5.1	-10.0	-3.9	-1.0	-1.4
M	-5.5	0.3	0.4	-1.6	-16.5	-3.9	1.4	-0.3
L	2.9	0.4	1.4	0.0	-4.5	-3.3	7.5	-1.0
T	-1.4	-0.4	-0.6	-1.4	-10.8	-3.8	1.2	-3.3

over the ensemble of eight forecasts in each month using the 30S-90N RAOB data as a reference for the verifications. The verifications were computed on model sigma layers, and then the RMSEs were computed using the verifications for all sigma layers within each cloud deck. The resulting deck RMSEs were squared, weighted by their deck thicknesses, and averaged and then the square root taken to get the "total" deck RMSEs. In each month, we computed DRMSE = 100 X [RMSE(PL-94) - RMSE(PL-92)]/RMSE(PL92) to arrive at the values shown in Table 5. Negative values reflect a decrease in RMSE in PL-94 compared to PL-92. The percent change of cloud amount RMSE was computed from the appropriate RMSE values in Table 4. Note that these values are derived from the 7-day verification period (days 18-24 of each month) while the T, V, and RH values are based on the eight forecasts in the ensemble, initialized on days spaced 3.5 days apart through the month.

The change in cloud amount RMSE due to the forecast model differences is small in both months. This is true in spite of the fact that RMSE decrease in T and to a lesser extent in V is larger in the July forecasts. The fact that there is no appreciable improvement in RH RMSE skill in either month may have contributed to the minimal changes seen in cloud amount RMSE. We say this because we found in CldAmt94 that RH and its variations were frequently leading predictors in the MLR diagnosis of cloud amount. Unfortunately, there is insufficient evidence in these results to arrive at any conclusions as to the impact of a change in NWP model forecast skill on cloud amount diagnosis skill. It may be necessary to realize larger differences in NWP forecast skill, particularly in relative humidity, before corresponding differences in cloud amount diagnosis skill are evident. We used the PL-94 forecast predictors for the cloud diagnosis discussed in the balance of this report.

6. ALTERNATIVE CLOUD AMOUNT DIAGNOSIS METHODS

In the previous study, we used multiple linear regression as the specific statistical forecast method, and model output statistics as the general statistical

forecast method. We found that, although we adjusted the regression slope to mitigate the effect, the MLR approach yielded too many cases of partial cloudiness and too few cases of clear and overcast compared to the transformed RTNEPH cloud distribution. In minimizing the least-squared error, the MLR arrived at less sharply defined cloud scenes than were apparent in the reference transformed RTNEPH depictions. To its credit, the MLR scheme did produce the lowest RMSE and MAE of any of the diagnosis methods tried in the previous study, but suffered in 20/20 score and normalized sharpness. We seek a method of relating the cloud amount predictand to the NWP forecast variable predictors that will maintain or reduce MAE and RMSE while at the same time produce cloud diagnoses that more closely resemble the cloud amount frequency distribution of the reference cloud data.

6.1 MLR Augmented by Regression Estimation of Event Probabilities

One of the principal findings of the previous study was a lack of sharpness of MLR relative to the sharpness of the transformed RTNEPH cloud amount distribution. We define sharpness as the number of grid points having cloud amounts in the range $0 \leq CA \leq 20$ percent and in the range $80 \leq CA \leq 100$ percent, divided by the total number of grid points having available cloud amount reports. As can be seen in Figure 2, the frequency of occurrence of transformed RTNEPH cloud amounts over the months of January and July 1991 in the Northern Hemisphere shows maxima at 0 and 100 percent, and values in between that are relatively similar to each other. MLR tended to skew the frequency distribution of cloud amount toward a distribution more like a normal distribution, with a maximum at about 50 percent cloud cover.

As a way of addressing this problem, we developed a "hybrid" MLR technique. We first developed a regression equation based on assigning a predictand value of 0 to clear gridpoints and a value of 1 to not-clear gridpoints. We then developed a

regression equation based on assigning a predictand value of 0 to not-overcast gridpoints and a value of 1 to overcast gridpoints. Thus, the predictand was assigned categorical values which were binary: either clear or not-clear on one hand, and not-overcast or overcast on the other. Using binary predictands as input to the multiple linear regression scheme is a two-category case of the use of multiple category predictands. The linear regression problems involves the development of a separate regression equation for each category. The solution of these equations by applying them to the predictor values at a gridpoint gives an estimate of the probability of each category being the correct one. Thus, this method of applying linear regression to categorical predictands has been called regression estimation of event probabilities[18,19] (REEP). We used binary REEP to compute the probability of not-clear at each gridpoint, and separately the probability of overcast at each gridpoint. We then applied the MLR scheme as usual to the percent cloud amount predictands. In diagnosing the cloud amounts at the forecast times we applied the resulting regression equations in order. We identified those gridpoints found to have low probabilities of not-clear (thus high probabilities of clear) first, then those gridpoints which had high probabilities of overcast next, then the remaining gridpoints were assigned the cloud amount as diagnosed using the MLR.

The hybrid REEP/MLR scheme we developed and used in this project had two versions. In version A, we used all of the randomly selected gridpoints' values of predictand and predictors to develop all three of the sets of regression equations: REEP clear/not-clear, REEP not-overcast/overcast, and MLR cloud amounts. That is, all development gridpoints in the 10-day development period were available to all three regression processes. In version B, we applied the procedures in sequence to the development gridpoint values, eliminating those gridpoints that very likely belonged to each group before applying the next procedure. In this way, each subsequent procedure was applied to only those gridpoints not likely to be identified with the previous procedure.

The first step in REEP/MLR-B was to develop the regression equation coefficients for clear/not-clear predictand values over all randomly selected gridpoints in the 10-day development period. The resulting regression was applied to all of the gridpoints that were used to develop it so that the probability of not-clear was diagnosed (the probability of clear was taken as 1-minus-probability of not-clear). We then sorted the probabilities of not-clear in increasing order. We also counted the number of reported clear gridpoints. The threshold probability (probability of not-clear below which a gridpoint is considered clear) was then set so that exactly the number of reported clear gridpoints would be identified. We then identified those gridpoints with not-clear probabilities below this threshold to be clear, and eliminated them from the equation development gridpoint pool for the not-overcast/overcast REEP and the cloud amount MLR. The not-overcast/ overcast REEP was then carried out on the remaining dependent predictor gridpoints, and the number of transformed RTNEPH overcast gridpoints was used to set the threshold probability using the resulting sorted probabilities. The gridpoints having a diagnosed probability of overcast greater then the threshold probability were identified and eliminated from the gridpoint pool. After elimination by both clear/not-clear and not-overcast/overcast regressions, the remaining dependent gridpoints were used for the development of the cloud amount MLR regression.

In the application of all three sets of regression coefficients to the independent forecast-time data, we followed the following sequence for both methods A and B. Each method is applied separately for each forecast time initialized from day 11, the day following the development period. First, the probabilities of not-clear and overcast are diagnosed from the respective REEP regression at each hemispheric gridpoint using the forecast-time predictor values at each gridpoint. Having counted the number of clear and overcast gridpoints and the number of all gridpoints having non-missing cloud amount values in the 10-day development period, we compute the frequency of occurrence of clear and overcast by dividing each total by the number of all gridpoints having non-missing cloud amount values. We then multiply the frequency of occurrence for both clear and overcast by the

42

number of hemispheric gridpoints to obtain an estimate of the number of gridpoints that should be in each of the two categories at the forecast time. We sort the diagnosed not-clear and overcast probabilities and identify the threshold probability whereby the estimated correct number of clear and overcast gridpoints will be identified. Any gridpoints not so identified are evaluated using the MLR cloud amount regression. Next, the number of gridpoints diagnosed by MLR to be clear or overcast (that is, have cloud amounts of 0 or 100 percent respectively) are counted. The number of MLR-clear gridpoints is subtracted from the number of clear identified by REEP and the threshold probability. This effectively decreases the probability threshold below which REEP clear/not-clear diagnoses clear. The number of MLR-overcast gridpoints is similarly subtracted from the number of overcast identified by REEP and the threshold probability. This effectively increases the probability threshold above which REEP not-overcast/overcast diagnoses overcast. Then the gridpoints having REEP not-clear probabilities less than the modified clear/not-clear threshold are assigned a diagnosed cloud amount of 0 percent. Any gridpoint failing that check has its overcast probability compared with the modified not-overcast/overcast threshold, and is assigned a 100 percent cloud amount if it is greater than the threshold. Any gridpoint that fails both checks is diagnosed using MLR and is assigned the corresponding cloud amount.

The design of the hybrid REEP/MLR cloud amount diagnosis method was intended to insure that the frequency of occurrence of the groups clear and overcast in the diagnosed cloud amounts at each forecast-time on day 11 were equal to the frequencies of occurrence of the corresponding groups in the 10-day development period. Of course, this does not insure a sharpness equal to that of the transformed RTNEPH, because cloud amount categories 5-20 percent and 80-95 percent are not forced to match RTNEPH frequency of occurrence values. In fact, cloud amounts in these ranges are the result of MLR in the hybrid method, so we would expect them to remain low in frequency of occurrence compared to transformed RTNEPH. Thus, sharpness should be improved in the REEP/MLR hybrid methods as compared with MLR, but will still not be as high as that of transformed RTNEPH.

43

Correspondingly, the RMSEs are bound to be larger in the REEP/MLR compared to MLR, because mean-squared error over all gridpoints is not allowed to be minimized--some gridpoints are being forced to have 0 or 100 percent cloud amount values. The goal of the REEP/MLR method is to maximize the growth of sharpness while minimizing the increase of RMSE.

6.2 REEP/MLR vs. MLR: Statistical Results

We diagnosed cloud amounts on the RTNEPH transform grid for the 7-day verification periods in January and July 1991 using MLR, REEP/MLR-A, and REEP/MLR-B. In all cases, the regression equation development period for each verification day was the 10 days prior to the day of diagnosis. Thus, for example, we used the 48-hour forecasts initialized on 8-17 January 1991 at 0000 and 1200 UTC to form the regression equation that was applied to the predictor values from the 48-hour forecasts initialized at 0000 and 1200 UTC 18 January 1991. In MLR, a single regression equation is developed and applied for the entire Northern Hemisphere in each cloud deck at each forecast duration. In the REEP/MLR methods, three regression equations (REEP clear/not-clear, REEP not-overcast/overcast, and MLR cloud amount) are developed and applied hemispherically for each deck and forecast duration.

In Table 6 we present the results of our cloud diagnoses as verified against the (timely) transformed RTNEPH cloud amounts at the forecast valid times. We computed the verification scores separately for the regions 0-30N, 30-60N, 60-90N, and 0-90N. However, our experience indicates that the errors are only truly minimized if the verification is conducted over the same area as was used for the statistical development. Thus, we show the verification scores only for the entire Northern Hemisphere in Table 6.

The bias values shown in Table 6 are acceptably small for all methods, cloud decks, and forecast durations. This is a known strength of MLR, and is proven again in these results. In the REEP/MLR methods, we forced the frequency of

Table 6. Cloud Amount Prediction Verification Scores for MLR and REEP/MLR-A, B Diagnosis Methods from PL-94 Forecasts for the Verification Periods 18-24 January and July 1991, Northern Hemisphere.

Deck	Forecast Hour	January			July		
		MLR	RMLR-A	RMLR-B	MLR	RMLR-A	RMLR-B

Bias (% Cloud Amount)

Deck	Forecast Hour	MLR	RMLR-A	RMLR-B	MLR	RMLR-A	RMLR-B
H	12	3.2	0.3	-2.4	5.3	-2.2	-1.4
	24	3.3	-5.1	-2.4	5.1	1.2	-1.8
	36	3.0	-5.3	-2.6	4.9	0.5	-1.9
	48	2.1	-5.6	-3.9	4.7	0.7	-2.0
M	12	1.0	-1.6	-3.4	3.3	0.7	1.6
	24	1.1	-0.6	-4.0	3.0	0.5	2.0
	36	0.8	-1.3	-4.1	2.9	0.5	1.4
	48	0.8	-2.3	-4.1	2.8	0.7	1.7
L	12	1.3	-1.1	-3.7	2.3	0.6	2.2
	24	1.2	-1.3	-2.7	1.9	0.6	4.5
	36	1.5	-1.3	-3.0	2.2	-0.3	3.2
	48	1.1	-1.5	-2.7	2.2	0.6	4.0
T	12	-3.4	-2.8	-3.7	-1.3	-1.6	2.8
	24	-3.2	-2.8	-2.4	-1.9	-1.7	2.5
	36	-2.5	-2.4	-1.8	-1.1	-1.2	1.0
	48	-3.0	-2.7	-1.2	-0.9	-0.9	2.6

RMSE (% Cloud Amount)

Deck	Forecast Hour	MLR	RMLR-A	RMLR-B	MLR	RMLR-A	RMLR-B
H	12	24.3	25.2	24.9	28.3	30.7	28.1
	24	24.7	28.1	25.6	28.7	29.7	28.7
	36	24.6	28.4	25.8	29.0	30.4	29.5
	48	26.8	28.7	25.4	29.5	30.9	30.1
M	12	32.6	34.4	34.5	34.2	36.8	37.7
	24	32.8	34.1	34.9	34.2	36.7	38.4
	36	32.9	34.7	34.6	34.6	37.2	38.5
	48	33.4	35.2	34.9	34.6	37.4	38.9
L	12	32.9	35.2	34.5	34.3	36.5	37.5
	24	33.2	35.4	35.6	34.5	36.3	38.4
	36	33.5	35.8	35.5	34.4	36.6	38.0
	48	34.2	36.0	35.9	34.6	36.7	38.7
T	12	34.0	35.3	38.9	33.4	34.2	37.2
	24	34.1	35.8	37.9	33.6	34.3	38.2
	36	34.8	36.4	38.6	33.9	34.7	36.5
	48	36.0	36.7	38.2	34.5	35.6	39.5

Table 6. Cloud Amount Prediction Verification Scores for MLR and REEP/MLR-A, B Diagnosis Methods from PL-94 Forecasts for the Verification Periods 18-24 January and July 1991, Northern Hemisphere.

Deck	Forecast Hour	January			July		
		MLR	RMLR-A	RMLR-B	MLR	RMLR-A	RMLR-B
				MAE (% Cloud Amount)			
H	12	14.6	13.4	12.6	18.0	17.5	16.9
	24	15.0	14.4	13.1	18.4	17.4	16.3
	36	15.2	14.7	13.2	18.7	17.9	16.9
	48	16.7	15.0	13.0	19.2	18.4	17.3
M	12	24.3	24.9	24.6	26.2	27.4	28.2
	24	24.6	24.8	24.9	26.5	27.6	28.8
	36	24.6	25.0	24.7	26.8	28.0	29.0
	48	25.2	25.5	25.0	27.1	28.2	29.5
L	12	24.9	25.9	25.1	26.9	28.0	29.0
	24	25.2	26.1	26.0	27.1	28.2	29.3
	36	25.2	26.6	26.0	27.1	28.3	29.4
	48	26.2	26.9	26.3	27.4	28.4	29.7
T	12	24.7	25.1	27.4	23.2	23.5	25.5
	24	25.0	25.7	26.8	23.7	23.8	26.5
	36	25.6	26.1	27.4	24.1	24.2	25.2
	48	26.8	26.4	27.2	24.7	25.0	27.8
				20/20 Score			
H	12	0.75	0.76	0.78	0.69	0.70	0.72
	24	0.74	0.77	0.77	0.68	0.68	0.72
	36	0.74	0.77	0.77	0.67	0.68	0.71
	48	0.72	0.77	0.78	0.67	0.67	0.70
M	12	0.56	0.55	0.57	0.51	0.50	0.49
	24	0.55	0.55	0.56	0.50	0.50	0.48
	36	0.55	0.55	0.56	0.50	0.49	0.48
	48	0.54	0.54	0.56	0.49	0.49	0.46
L	12	0.53	0.52	0.55	0.49	0.48	0.47
	24	0.53	0.52	0.53	0.49	0.48	0.47
	36	0.52	0.51	0.53	0.49	0.48	0.46
	48	0.50	0.50	0.52	0.48	0.47	0.47
T	12	0.56	0.55	0.54	0.59	0.59	0.57
	24	0.55	0.54	0.54	0.58	0.58	0.56
	36	0.54	0.54	0.54	0.57	0.57	0.57
	48	0.52	0.53	0.53	0.56	0.55	0.54

Table 6. Cloud Amount Prediction Verification Scores for MLR and REEP/MLR-A, B Diagnosis Methods from PL-94 Forecasts for the Verification Periods 18-24 January and July 1991, Northern Hemisphere.

Deck	Forecast Hour	January			July		
		MLR	RMLR-A	RMLR-B	MLR	RMLR-A	RMLR-B
				Normalized Sharpness			
H	12	0.87	0.92	0.97	0.84	0.95	0.95
	24	0.86	1.00	0.96	0.83	0.90	0.95
	36	0.86	1.00	0.97	0.82	0.91	0.95
	48	0.87	1.02	0.97	0.82	0.90	0.95
M	12	0.80	0.88	0.92	0.68	0.78	0.72
	24	0.79	0.86	0.94	0.64	0.76	0.72
	36	0.80	0.88	0.92	0.63	0.76	0.71
	48	0.79	0.88	0.93	0.58	0.74	0.67
L	12	0.72	0.81	0.83	0.59	0.68	0.69
	24	0.71	0.80	0.82	0.60	0.69	0.76
	36	0.70	0.79	0.79	0.58	0.67	0.70
	48	0.67	0.78	0.80	0.57	0.67	0.76
T	12	0.77	0.82	0.91	0.84	0.85	0.96
	24	0.74	0.80	0.88	0.81	0.83	0.96
	36	0.75	0.81	0.88	0.80	0.82	0.90
	48	0.73	0.80	0.85	0.78	0.80	0.94

occurrence of gridpoints with 0 percent and 100 percent cloud amounts at the diagnosis times to be the same as the frequency of occurrence of 0 percent and 100percent respectively in the 10-day development period. Thus, it is expected that no appreciable bias would be imposed because the frequency of occurrence for the 10-day development period should be representative of the frequency of occurrence on the diagnosis day (day 11). We note that the magnitudes and signs of the biases are very consistent for a given method and cloud deck. There appears to be no noticeable growth in bias with forecast duration for any of the methods.

The RMSE scores in Table 6 do show discernible trends in method and forecast duration. A common feature among all methods is the small increase in RMSE with forecast duration. As expected, the MLR scores are better than their REEP/MLR counterparts in every cloud deck/method entry except two. The better REEP/MLR RMSEs vary with cloud deck and month. REEP/MLR-B is better for January/high clouds, RMSE/MLR-A is better for January/total clouds, and the results are mixed for January/middle, low clouds. REEP/MLR-B has lower RMSEs for July/high clouds, while REEP/MLR-A has lower RMEs for July/middle, low, and total clouds.

When we looked at the regional verification scores, we found that the REEP/MLR-B schemes tend to produce large values of positive cloud amount bias, and consequent large RMSE and MAE, in the 60-90N region for January/total cloud and July/middle, low, and total cloud. In the deck/month sections of the table not affected by this problem, REEP/MLR-B is either clearly or marginally better than REEP/MLR-A. This suggests that REEP/MLR-B is potentially a better method than REEP/MLR-A. It would be necessary to execute the regional version of REEP/MLR-B and verify the results to validate the benefit of the withholding of identified clear and overcast gridpoints in the subsequent regressions as done in REEP/MLR-B.

We see that the REEP/MLR-B MAE values are lower than those of REEP/MLR-A in virtually every entry of the cloud deck/month sections of the table not affected by the large 60-90N cloud amount bias of REEP/MLR-B. These sections are:

48

January/high, middle, and low clouds, and July/high clouds. In fact, in the high clouds in both months, the REEP/MLR MAEs for high clouds are better than the MLR MAEs, especially for REEP/MLR-B. Looking back at Figure 2, we see the reason for this. Some 60-70 percent of the timely high cloud gridpoints are clear. Using the REEP/MLR methods to ensure the proper number of clear gridpoints benefits high cloud more than the other decks because of the dominant number of clear high cloud cases. In any case, the MAE scores lend support to the perception that REEP/MLR-B may be better than REEP/MLR-A when both are executed regionally.

The first observation made when looking at the 20/20 scores in Table 6 is the degree of similarity of the values among the three methods, particularly in middle, low, and total clouds. The 20/20 score is reduced very little (less than 10 percent) in all cases with increasing forecast duration. Again, REEP/MLR-B scores best in the high cloud deck in both months. REEP/MLR-B is as good as or better than the other two methods in January/middle and low decks as well. It is only in the deck/month sections affected by the 60-90N cloud amount bias of REEP/MLR-B that the latter method results in worse 20/20 scores. This further lends support for the value of the REEP/MLR-B methodology. Again, this extra benefit would have to be validated by a regional execution of the methods.

Finally, as we expected, the REEP/MLR methods do produce improved normalized sharpness over MLR. From the normalized sharpness results displayed in Table 6 we see that both REEP/MLR methods produce a greater percentage of gridpoints with cloud amounts less than or equal to 20 percent and greater than or equal to 80 percent. The degree of improvement of sharpness over MLR varies with deck, month, and REEP/MLR method. Method A appears to exhibit greater sharpness than method B in January/high clouds (except at 12 hours) and in July/middle clouds. Method B is sharper in all other deck/month sections of the table, but only marginally so in January/low clouds. In all methods, there is no clear trend for sharpness to decrease with increasing forecast duration.

49

It is of interest to determine how the frequency distribution of cloud amount was changed by the REEP/MLR methods to arrive at this modest increase in normalized sharpness over MLR. In the REEP/MLR methods, we attempted to preserve the frequency of occurrence of only 0 percent and 100 percent cloud amount categories. The other 5 percent cloud amount categories in the definition of sharpness (5-20 percent and 80-95 percent) were not included in the REEP process and were thus subject to the MLR. It is of interest to determine which 5 percent cloud amount categories had reduced frequencies of occurrence in order boost the frequency of occurrence of 0 percent and 100 percent cloud amount categories in the REEP/MLR methods. With this in mind, we present the frequency distribution plots shown in Figure 13. The results show that the frequency of occurrence was reduced from MLR values by the REEP/MLR in all other five-percent categories to increase the 0 and 100 percent frequency of occurrence. The reduction is generally greater in method A in the 5-20 percent and 80-95 percent categories, and is greater in method B in the intermediate (25-75 percent) 5 percent categories. This would explain why the normalized sharpness is generally larger for REEP/MLR-B than for REEP/MLR-A in the 12-hour forecasts.

The distribution of the reduction of frequency of occurrence by the REEP/MLR methods results in a frequency distribution that agrees better with RTNEPH than does MLR at 0, 100, and 25-75 percent cloud amount, but worse than MLR in the 5-20 and 80-95 percent ranges. The greater reduction of frequency of occurrence in these latter ranges results from the fact that these are the gridpoints most likely to be clear or overcast in the probability-based REEP schemes. Thus, these gridpoints with what MLR would assign as near-clear or near-overcast are more likely to be selected by the REEP scheme as clear or overcast to make its quota of such points. If one were to broaden the range of the clear and overcast categories to include more than just 0 and 100 percent respectively, then the category as a whole would be boosted from its MLR values rather than robbing from near-clear and near-overcast to pay clear and overcast.

50
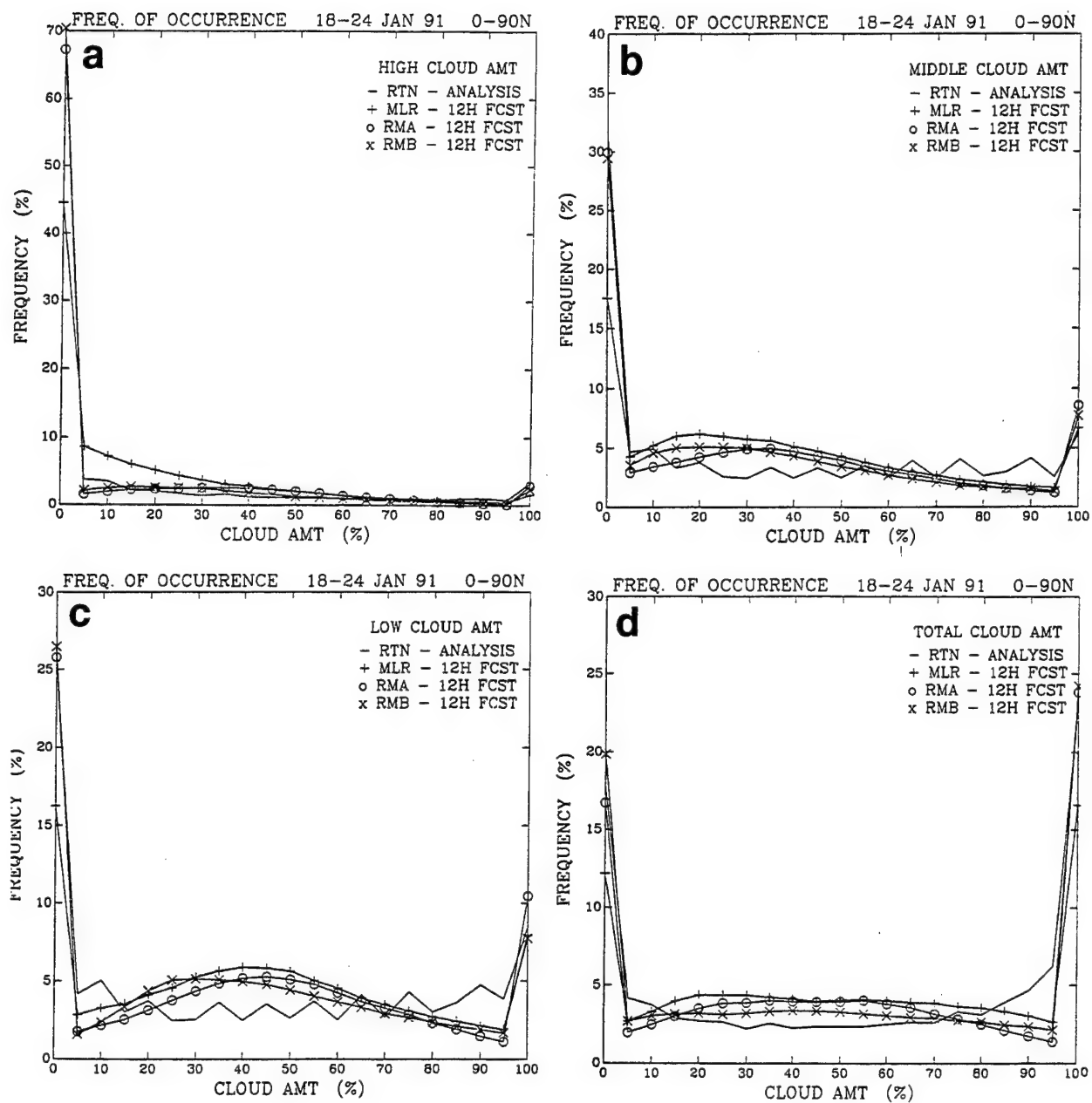
Figure 13. Frequency of occurrence of Cloud Amount, Twice-Daily During the January Verification Period of RTNEPH (RTN), MLR, REEP/MLR-A (RMA), and REEP/MLR-B (RMB) for (a) High, (b) Middle, (c) Low, and (d) Total Cloud Decks.

51

As can be seen from Figure 13, method B does not necessarily always result in larger 0 percent or 100 percent frequency of occurrence than method A (in four of the eight 0 percent and 100 percent values method B frequency of occurrence is lower). However, the frequency of occurrence increase in clear and overcast appears to be drawn more from the intermediate 5 percent categories, and thus sharpness is better than for method A.

We mentioned earlier that the primary goal of the REEP/MLR methods was to maximize the increase of sharpness while minimizing the increase in RMSE with respect to MLR. Table 7 shows quantitatively how well the two REEP/MLR methods accomplished this goal. For the 12-hour forecast cloud diagnosis results shown in Table 6, we computed the quantity

$$\frac{[NS(REEP/MLR)-NS(MLR)]/NS(MLR)}{[RMSE(REEP/MLR)-RMS(MLR)]/RMSE(MLR)}$$

for all deck/month entries. In all but two cases, the magnitude of the ratio is greater than one, showing that on a percentage basis the sharpness gains more than the RMSE loses. (The negative value in July/high cloud/method B is due to the fact that RMSE for REEP/MLR-B is smaller than for MLR). We might expect that the values of the four entries affected by the 60-90N high positive bias in REEP/MLR-B might be higher if the regressions were computed and applied regionally.

To summarize these results, we note that the REEP/MLR methods generally succeed in improving the frequency distribution of cloud amounts compared to MLR. By better preserving the maxima of frequency of occurrence of 0 and 100 percent cloud amount, the frequency of occurrence in the 25-75 percent cloud amount categories is decreased, which in turn generally leads to somewhat increased RMSEs compared to MLR. In MAE and 20/20, the REEP/MLR methods produce scores that are close in magnitude to the MLR scores. Apart from the apparent problem that REEP/MLR-B has of excluding a preponderance of either clear or cloudy points from a certain latitude band from the development of one of

its regressions, method B appears to have overall cloud diagnosis skill advantages over REEP/MLR-A.

Table 7. Ratio of the Percent Normalized Sharpness Change to the Percent RMSE Change for REEP/MLR Methods, 12-h Forecasts, 18-24 January, July 1991, Northern Hemisphere.

	January		July	
Deck	RMLR-A	RMLR-B	RMLR-A	RMLR-B
H	1.6	4.7	1.5	-18.5
M	3.1	2.6	1.9	0.6
L	1.8	3.1	2.4	1.8
T	1.7	1.3	0.5	1.3

6.3 REEP/MLR vs. MLR: Cloud Amount Map Comparisons

We now consider a few examples of cloud amount distribution maps resulting from the MLR and REEP/MLR cloud amount diagnosis methods. It is important for the user to be able to see realistic depictions of future cloud scenes over his/her area of interest, for use in forecast guidance. The following cases were chosen from the 7-day verification periods to visually represent the performance of the cloud diagnosis methods.

6.3.1 0000 UTC 25 JANUARY 1991 EUROPE CASE

Figure 14 displays the high cloud deck distribution over most of Europe on 0000 UTC 25 January 1991. We have shown the transformed RTNEPH and the 12-hour forecast valid at this time and date produced by the three cloud amount diagnosis methods: MLR, REEP/MLR-A, and REEP/MLR-B. The RTNEPH depicts a band of high clouds over northern and eastern Scandinavia, and a small patch of high cloud on the Spanish Mediterranean coast. The three diagnosis methods all produce some cloudiness over Scandinavia but none of them produce the > 80 percent cloud band seen in the RTNEPH. The REEP/MLR-A comes closest, with perhaps a small blob of > 80 percent over the northwestern coast of Norway. REEP/MLR-B is the

Figure 14. High Deck Cloud Amounts (%) Over Europe on 0000 UTC 25 January 1991 from (a) Transformed RTNEPH, (b) 12-Hour MLR Forecast, (c) 12-Hour REEP/MLR-A Forecast, (d) 12-Hour REEP/MLR-B Forecast.

54

diagnostic method that most underestimates the Scandinavian cloudiness. All three diagnostic schemes produce a band of apparently spurious cloudiness northeast of the Black Sea, where the RTNEPH shows just a hint of cloudiness. Allthree methods also attempt to represent some cloudiness in the western Mediterranean Sea, which does appear on the RTNEPH depiction.

The RTNEPH cloudiness in the middle cloud deck depiction (Figure 15) exists in three areas: Scandinavia, north and south of the Black Sea, and the western Mediterranean. Central Europe appears to be cloud-free in this depiction. The three diagnosis methods maintain the cloudless central European area. They also do a good job of locating the boundary of the > 80 percent cloud mass in northern Scandinavia, and the < 20 percent boundary in southern Scandinavia. All three methods locate the cloud amount maxima over the Black Sea rather than to its north and south as in the RTNEPH depiction. All three methods depict lesser cloud amounts over the central Mediterranean and the northern Iberian peninsula which don't exist in the RTNEPH. All three methods represent the western Mediterranean cloudiness, with the erroneously largest amounts over the largest area produced by REEP/MLR-B.

In addition to the cloudy areas depicted on the middle cloud map, the RTNEPH low cloud map (Figure 16) shows a north-south band of moderate cloud amounts over central Europe and between the Black and Caspian Seas. All three diagnostic methods produced less than 20 percent cloudiness south of the Baltic Sea where RTNEPH shows moderate cloudiness. This cloudless area is largest in REEP/MLR-B. The methods also underestimate the cloud amounts farther south of the northern Balkan peninsula - again, REEP/MLR-B underestimates the most. All three methods do correctly produce clouds between the Black and Caspian Seas. In this case, REEP/MLR-B seems to have the best estimate of western Mediterranean cloudiness. Over the North Sea, RTNEPH shows a very cloudy low deck scenario. None of the methods produce enough cloudiness over the North Sea, although REEP/MLR-A does best (REEP/MLR-B most underestimates the cloudiness here).

In the total cloud amount maps (Figure 17), we see that the cloud diagnosis methods properly represent the > 80 percent cloud amounts over northern Scandinavia, but position its southern boundary too far south over the Baltic Sea, too far north over the North Sea, and in REEP/MLR-B too far south over northwestern Russia. All three methods produce a southwest-northeast swath of > 80 percent total cloud that extends over almost the entire Black Sea, whereas the RTNEPH ends the > 20 percent cloudiness on the northeast shore. All three methods do a reasonable job of locating the > 80 percent cloud cover over the western Mediterranean. However, all three cloud schemes produce entirely too little total cloudiness over central Europe (between the Baltic and Adriatic Seas). This is especially true of REEP/MLR-B, in which a huge area of < 20 percent clouds covers north-central Europe and the North Sea. The latter location is where RTNEPH depicts > 80 percent cloud amounts, as well as small patches of the same between the Baltic and Adriatic.

The sharpness of the REEP/MLR-B is more apparent in total cloud than in any of the decks. Notice how tight the gradients are between < 20 and > 80 percent cloud amounts areas. This is an asset of the REEP/MLR-B, but it is accomplished at the expense of loss of accuracy in some areas, especially central Europe. The figures in Table 7 suggested that REEP/MLR-B has at least the potential to increase sharpness more than REEP/MLR-A while maintaining at least as much accuracy. As we saw in the RMSE scores in Table 6, REEP/MLR-B January total cloud suffered a greater loss of accuracy than REEP/MLR-A. Perhaps the regional application of the methods would reverse this and benefit from the sharpness advantage of method B without the extent of accuracy loss.

Another point to make about the total cloud depictions is the fact that in some areas, (most notably central Europe), there is more low cloud than total cloud in the diagnoses. This apparent contradiction can be explained by the fact that the cloud diagnoses in each deck (including total cloud) were accomplished entirely separately from each other. Thus, the total cloud diagnosis development and
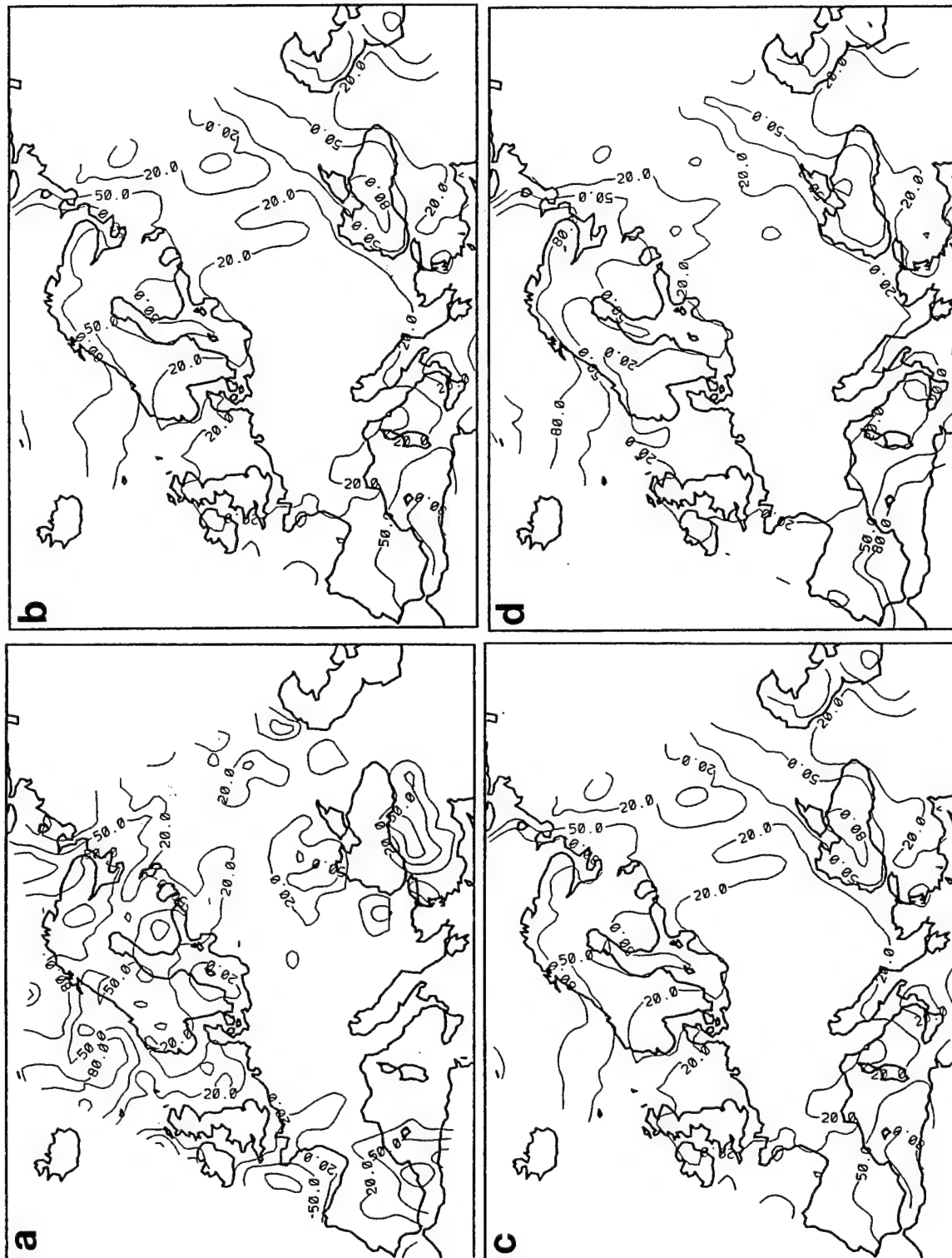
Figure 15. Same as in Figure 14 for Middle Deck Cloud Amounts.

57
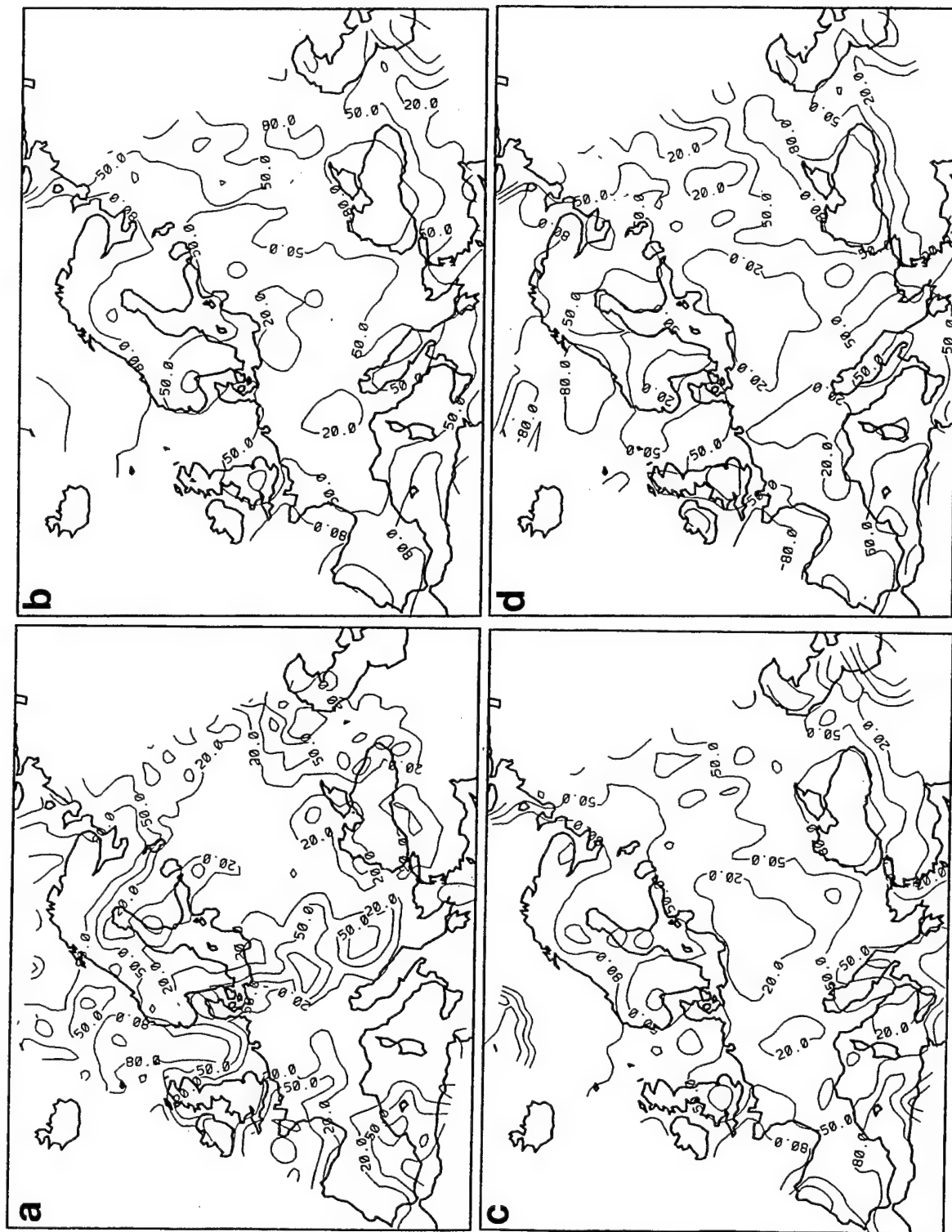
Figure 16. Same as in Figure 14 for Low Deck Cloud Amounts.

58
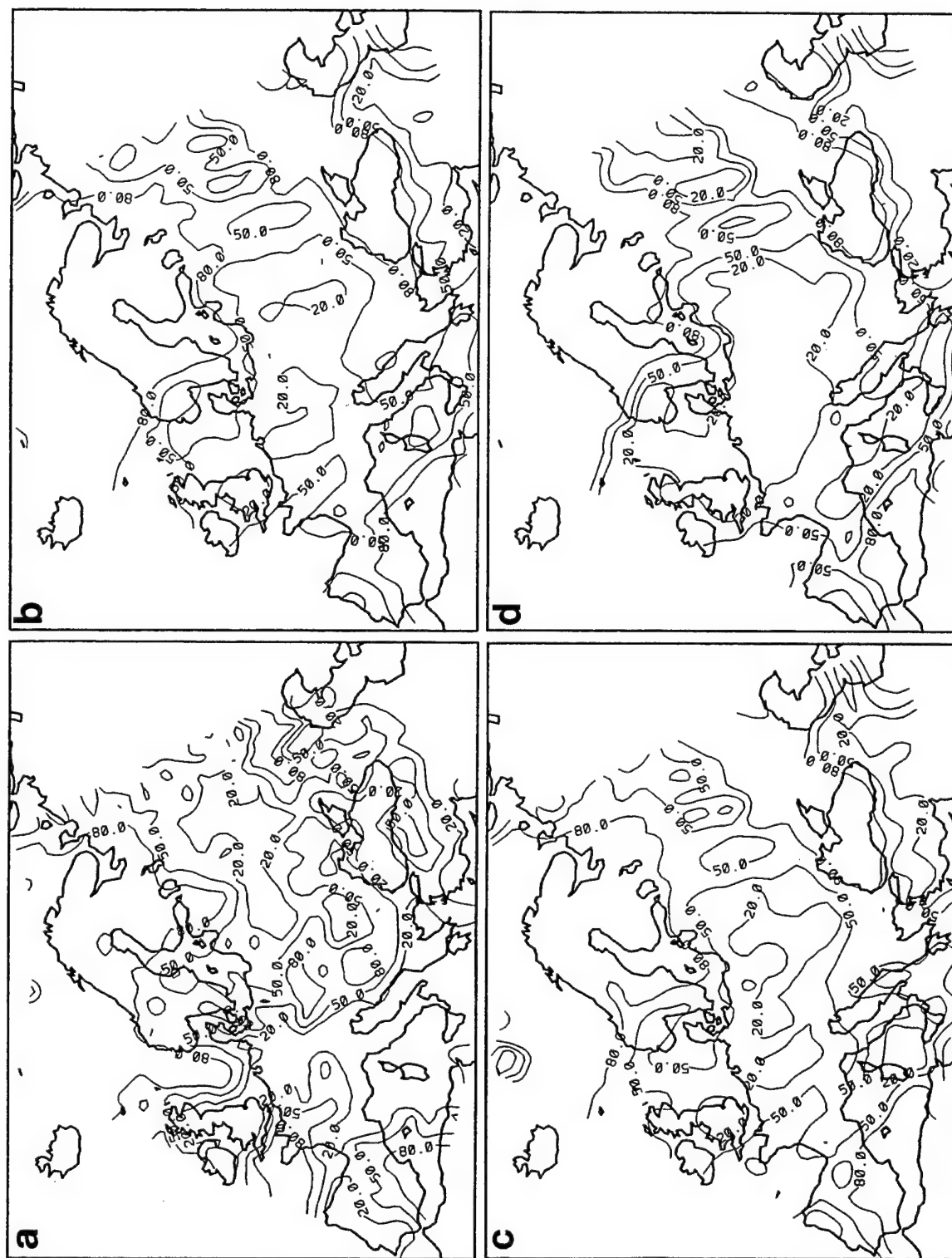
Figure 17. Same as in Figure 14 for Total Deck Cloud Amounts.

application had no knowledge of the individual deck cloud amounts. This approach has the advantage of giving an independent estimate of the overall cloud over a region, along with a measure of the uncertainty of the cloud diagnosis. Those areas in which we see greater disagreement between conglomerate deck cloud cover and the separately-derived total cloud cover would be deemed to have a higher degree of uncertainty in the forecast.

6.3.2 0000 UTC 25 JANUARY 1991 EASTERN ASIA CASE

In Figures 18-21 we show a case of cloudiness over eastern Asia that is much less complex than the European case just discussed. This case highlights the differences in the diagnosis products with more clarity, because it shows essentially distinct cloud bands over rather limited portions of the area shown.

In Figure 18, we have depicted the high cloud amounts for RTNEPH and for the 12-h forecast diagnoses from MLR and the REEP/MLR methods. The diagnostic methods produce cloudiness only in the east-central portion of the figure, where the RTNEPH shows the primary cloudy areas to be located. The predictors apparently indicate an area of > 50 percent cloudiness that is not shown in the RTNEPH. However, the primary area of cloudiness in eastern China is well represented by all three diagnostic methods.

The diagnosed middle deck cloud amounts shown in Figure 19 fare much worse in representing the actual cloud distribution. All three diagnostic methods completely miss the primary patch of cloudiness in central-to-eastern China, and they overdo the degree of cloudiness in the extreme western edge of the region, especially REEP/MLR-A. There is a strong similarity in the pattern of cloudiness between MLR and REEP/MLR-A--they are essentially the same except for the small areas of > 80 percent cloud amount on the western edge in REEP/MLR-A. In this case, REEP/MLR-B produces the least amount of clouds in the western area and none at all in central-eastern China. All three methods capture the partial cloudiness portrayed along the extreme eastern coast of China.

Figure 20 shows the low cloud distribution of the diagnosed cloud with reference to the RTNEPH. Like the high cloud but unlike the middle cloud, the major area of low cloud diagnosed is in the central and eastern portion of China. All three methods position this cloud area fairly well but underforecast the amounts. The area of > 80 percent is restricted to the ocean area off the coast in the methods but clearly overspreads SE China in the RTNEPH. Both MLR and REEP/MLR-A over-represent the western cloudiness, whereas REEP/MLR-B somewhat under-represents it.

A comparison of the maps of total cloud amount (Figure 21) with those of low cloud in Figure 20 illustrates the point made earlier about information on reliability. The major cloud mass is located in generally the same position for both low and total cloud for all three methods, which would tend to increase the probability of the future existence of cloudiness in this area in a operational forecast scenario. There are no obvious inconsistencies between low and total cloud amount distributions in the diagnoses. Even though the size of the area covered by > 80 percent cloud amount is too small in the diagnosis, this is probably a better diagnosis than was produced in the decks. Notice how all three methods fail to capture the sharp gradient in cloud amount surrounding the large cloud mass over SE China. We thought that by forcing the required frequency of occurrence of 0 percent and 100 percent cloud amount in the REEP augmentation of MLR, the gradients of MLR would be more realistic. It appears that the primary benefit occurred in the low end of the cloud amount spectrum--the 20 percent contours are smoother in the REEP methods, and there is better agreement of this boundary with RTNEPH. The fact that the high end (> 80 percent) values do not cover enough area in the cloud mass is the reason for lack of sharpness of gradients.

7. CLOUD AMOUNT CATEGORY DIAGNOSIS USING MULTIPLE DISCRIMINANT ANALYSIS

Although the REEP/MLR methods were successful in increasing the sharpness of the diagnosed cloud amount distribution, they still were based on linear
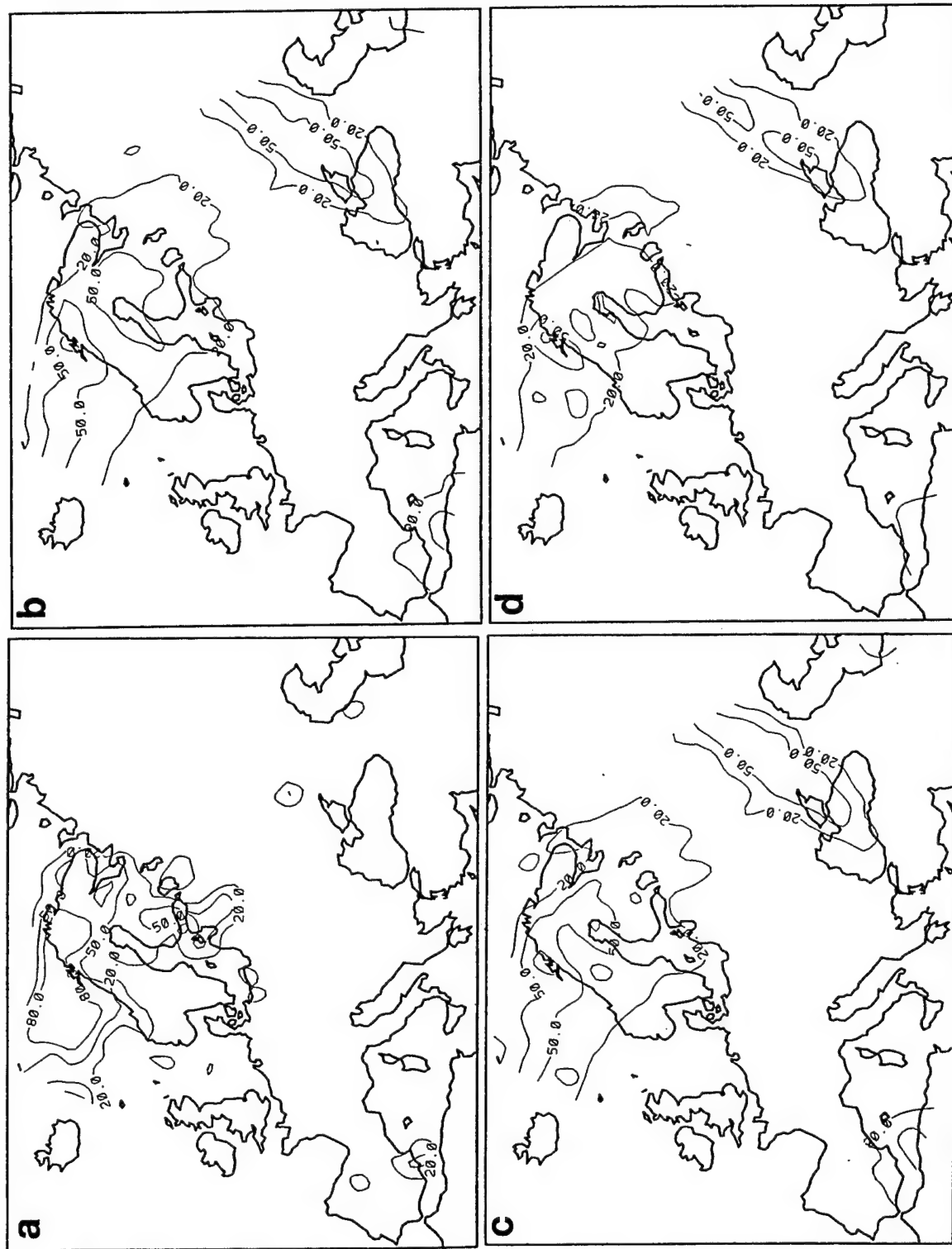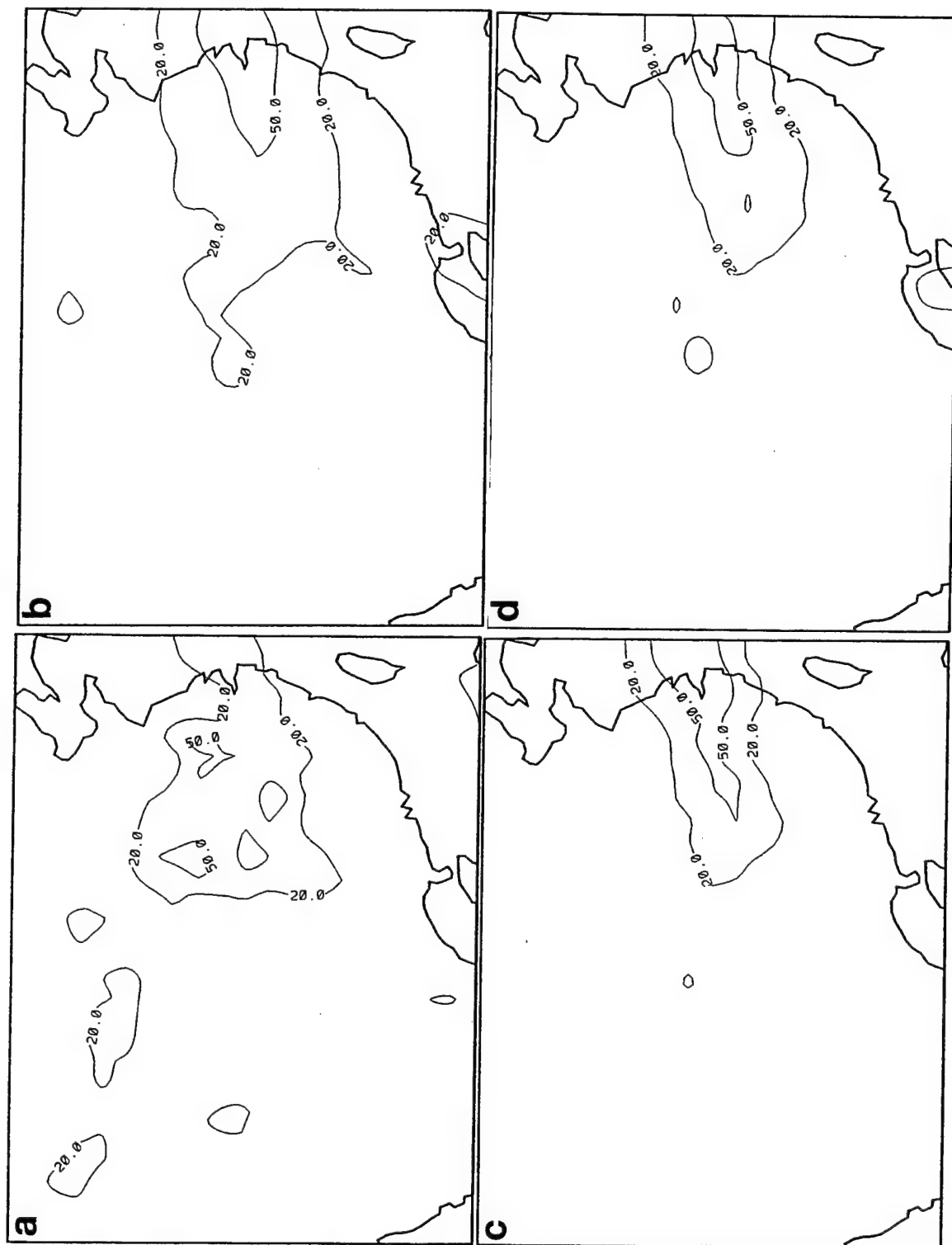
Figure 18. High Deck Cloud Amounts (%) Over Eastern Asia on 0000 UTC 25 January 1991 from (a) Transformed RTNEPH, (b) 12-Hour MLR Forecast, (c) 12-Hour REEP/MLR-A Forecast, (d) 12-Hour REEP/MLR-B Forecast.
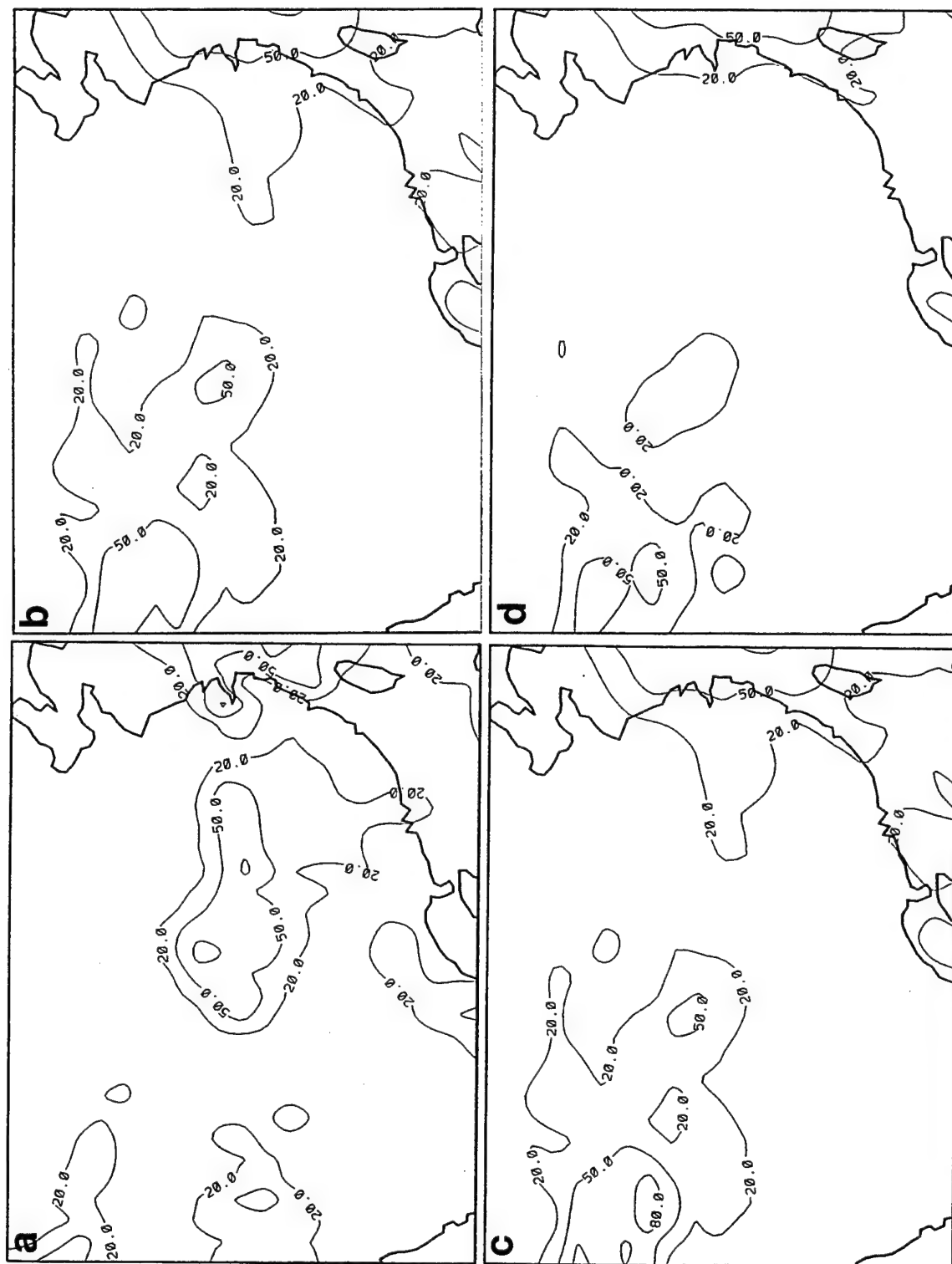
Figure 19. Same as in Figure 18 for Middle Deck Cloud Amounts.

63
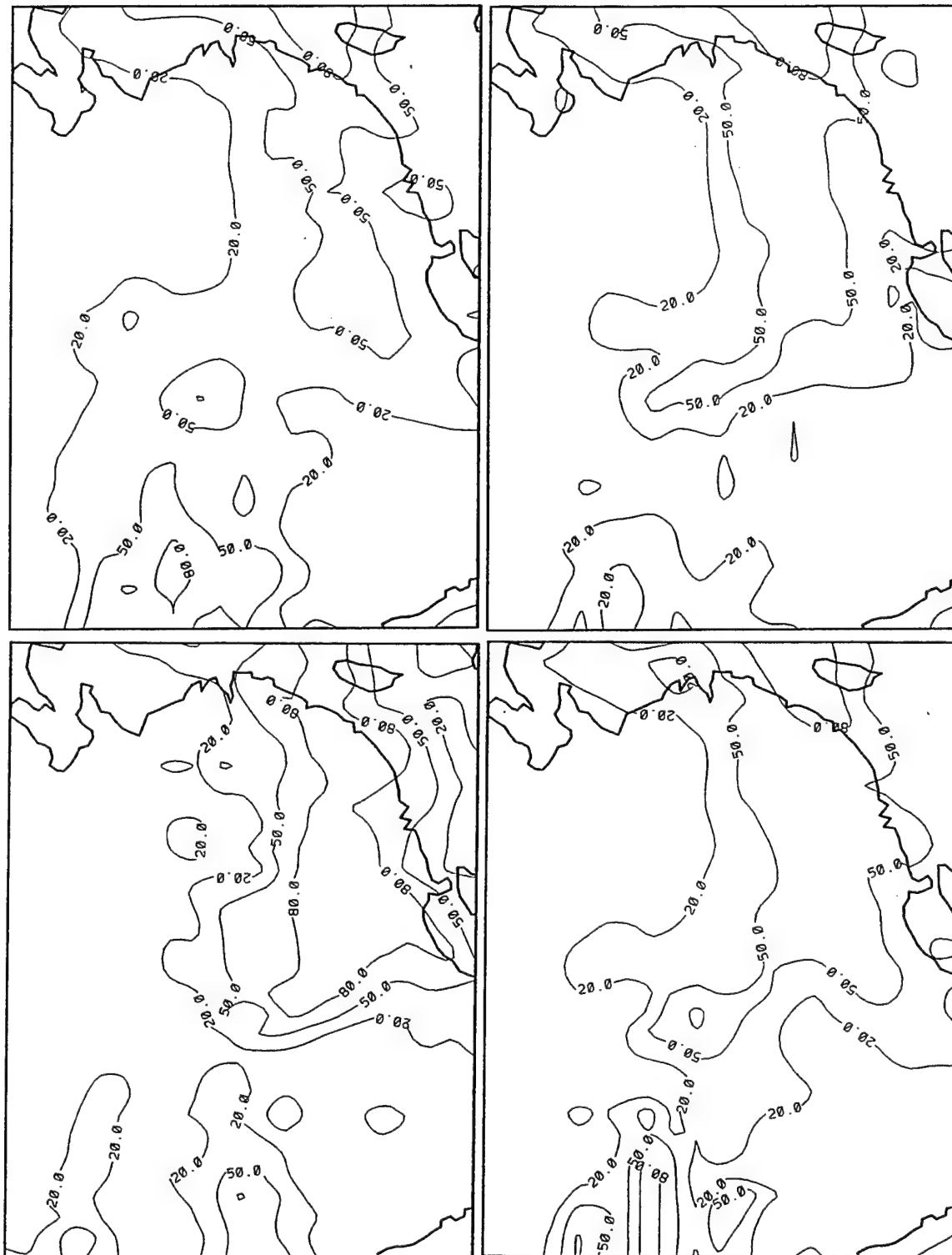
Figure 20. Same as in Figure 18 for Low Deck Cloud Amounts.

64
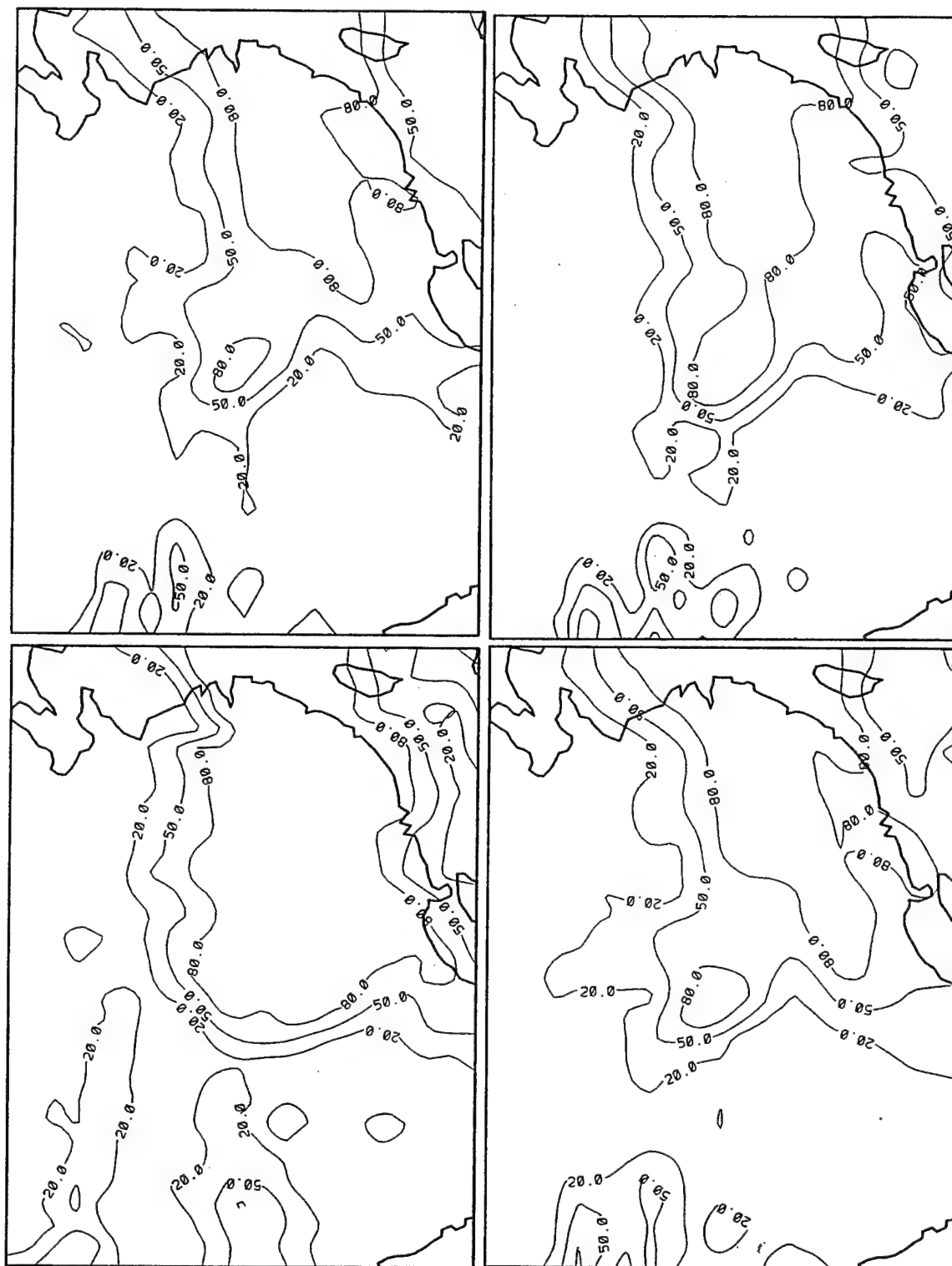
Figure 21. Same as in Figure 18 for Total Deck Cloud Amounts.

regression. The cloud maps we just discussed revealed many similarities between the diagnoses of cloud amount by MLR and the REEP/MLR methods, particularly REEP/MLR-A. REEP/MLR-B showed the potential to have the most benefit of the three methods in the verification statistics. However, treating 0 percent and 100percent cloud amounts as separate categories had the effect of reducing the frequency of occurrence of neighboring five-percent cloud amount categories (5-20 percent and 80-95 percent). If we could boost the frequency of occurrence of the 0 percent, 5-20 percent and 85-100 percent categories as a whole, we might be better served in reproducing the RTNEPH cloud amount frequency distribution.

With this goal in mind, we began considering the use of a categorical statistical diagnosis method. By this we mean a method in which the predictand is a group of five-percent cloud amount categories instead of treating the predictand as a continuous function ranging between 0 and 100 percent cloud amount. Our intuition tells us that the characteristics of the natural physical relationships between our pool of predictors and cloud amount do not allow for discerning unique categories of predictors that have a one-to-one relationship with corresponding five-percent categories of cloud amount. Therefore, any categorical diagnosis method would require that the five-percent categories be lumped together into a fewer number of more broad categories that collectively span the spectrum between 0 and 100 percent.

One approach to categorical diagnosis is to continue with regression. The basic difference between continuous predictand regression and categorical predictand regression is that the predictand in the latter can only take on discrete values. These would be category indices that each correspond to a range of cloud amounts in our case. Then for a particular event, all categories would have a value of zero except the category that contained the cloud amount actually observed, which would have a value of one. This then requires the development of a regression equation for each of N possible events; that is, the observed cloud amount could fall into N possible categories between 0 and 100 percent inclusive. Once the N regression equations are developed, they can be applied to an independent set of

66

predictor data and solved for the probability (between 0 and 1) of each of the N events being true. Although in the development data only one event is true and it is unambiguous, in the application data there are N possible events that could be true, and the probability of any of the events is almost never equal to 1. The probabilities of the N events must add up to 1, but there may be many cases in which the probabilities may be nearly equal to each other. It is difficult to choose which event is true in such cases, even though we know that only one event can be true. The method of developing the categorical regression equations is called regression estimation of event probabilities, or REEP.[18,19]

Another quite different approach to categorical diagnosis is called multiple discriminant analysis[20,21] (MDA). In MDA, the predictor values are used to discriminate between the categories of the predictand. The discriminant analysis is a procedure where relationships are sought that maximize this ability to distinguish between categories. Like REEP, its output is a set of probability forecasts of the predictand categories.

MDA is carried out in three steps. First, the method is applied to a dependent sample of predictor data and corresponding predictand category values to obtain a set of discriminant functions. These functions are linear combinations of the predictors that provide a basis for determining the probabilities of the event categories. Next, these discriminant functions are applied to independent predictor data to determine the probability of occurrence of each predictand category. Finally, it is necessary to use a method to select the most likely predictand category based on the calculated category probabilities.

7.1 Preliminary Development and Testing of Multiple Discriminant Analysis

The multiple discriminant analysis algorithm DSCRM of the IMSL STAT/LIBRARY was applied to 3 individual days of our 7-day verification period in both January and July (dates 18, 21, 24). We used the predictors from the 12-hour PL-94 forecasts for the 10-day periods prior to each of the dates, and the

corresponding transformed RTNEPH cloud amount predictands, to develop the discriminant functions to be applied to the succeeding day's predictors in each case.

The algorithm DSCRM does not have the facility to select the most discriminating predictors from among a longer list of available predictors. Instead, it simply uses the provided predictors, and attempts to discriminate between the "groups" of predictor values that correspond to discrete groupings of the predictand. In our case, we provided only the values for the 20 leading predictors for each 10-day period as selected by the MLR forward stepwise regression.

Before entering the algorithm, the continuous cloud amount (0-100 percent) predictand is assigned to discrete categories, or groups; that is, the predictand values are assigned a group index instead of a continuous cloud amount value. The algorithm DSCRM can then specify the extent to which the predictor values that correspond to the predictand groups are distinct for the various groups specified. This degree of distinctiveness of the predictors between the groups is measured by a quantity known as the Mahalanobis Distance, or the Mahalanobis D^2 Statistic. This quantity is the ratio of the difference between group means of any two groups of predictors to the variance of the predictors within the two groups. The larger the value of D^2 between any two groups of the predictors, the more likely that MDA will be able to assign the right predictand category to a specified vector of predictor values; that is, the less ambiguity there is between the groups of predictor values. Thus, an appropriate predictor selection method for MDA would involve a screening of predictors not by their total correlation with the predictand as in MLR, but on the basis of their collectively maximizing the value of D^2 between the various pairs of groups. The algorithm DSCRM does not appear to have a capability for predictor selection on this basis, and the development of such a screening procedure was outside the scope of this project. Further consideration of MDA as a tool for cloud diagnosis should include obtaining and using a predictor selection method based on maximization of group-group D^2.

The values of D^2 for the predictors provided are a direct output of the algorithm DSCRM. The group-group D^2 values for the 10-day development period 8-17 July

1991 are given for middle cloud in Table 8 and for total cloud in Table 9 (the trends for high and low cloud were the same as for middle and total cloud). The D^2 values are shown for four different 4-category groupings of the predictand cloud amount. The cloud amount ranges assigned to each grouping's category indices are given in Table 10. For example, in Grouping 1, 0 percent cloud amount was assigned a category value of 1, 5-40 percent assigned a value of 2, 45-80 percent assigned a 3, and 85-100 percent assigned a 4 before entering the MDA algorithm. Recall that the transformed RTNEPH consists of 21 categories, 5 percent cloud amount intervals from 0 to 100 percent inclusive. Because we were certain that the full 21 categories would not result in appreciable category distinctiveness, we chose 4-category groupings to insure sufficient distinctiveness. We see from Tables 8 and 9 that, for this 10-day period, Grouping 3 (the same as clear, scattered, broken, overcast) consistently demonstrates the highest D^2 values between the various groups of any of the groupings. This may suggest that there are enough distinctive weather features accompanying an observation of overcast to justify considering it as a category by itself, rather than including it in a category with subovercast conditions. Indeed, the trend in the table values shown is that the more 5 percent categories that are included with 100 percent cloud amount in category 4, the lower the D^2 values between category 4 and the other three categories become. The fact that the D^2 values for the other three category pairs is highest for Grouping 3 could be due to the fact that this grouping has the largest range of 5 percent cloud amount categories for categories 2 and 3.

After developing the discriminant functions for the three 10 periods 8-17, 11-20, and 14-23 July 1991 for the entire Northern Hemisphere, we applied them to the hemispheric predictor values at each equal-area gridpoint for the 12-hour forecasts initialized on the eleventh day in each case (18, 21, 24 July 1991). This application produces values for the probability of the diagnosed cloud amount falling within each of the designated categories. The degree to which one of the categories has a probability near to one (100 percent probability) is the degree to which the MDA is able to discriminate between the groups of predictor values corresponding to the

Table 8. Mahalanobis D² Statistic for 8-17 July 1991, Middle Cloud (Column and Row Labels are Group Numbers).

Grouping #1

	1	2	3	4
1	0.00	0.75	1.71	2.26
2	0.75	0.00	0.30	0.68
3	1.72	0.30	0.00	0.11
4	2.26	0.68	0.11	0.00

Grouping #2

	1	2	3	4
1	0.00	0.63	1.55	2.07
2	0.63	0.00	0.30	0.66
3	1.55	0.30	0.00	0.09
4	2.07	0.66	0.09	0.00

Grouping #3

	1	2	3	4
1	0.00	0.82	1.81	2.56
2	0.82	0.00	0.30	0.87
3	1.81	0.30	0.00	0.25
4	2.56	0.87	0.25	0.00

Grouping #4

	1	2	3	4
1	0.00	0.72	1.63	2.17
2	0.72	0.00	0.27	0.63
3	1.63	0.27	0.00	0.11
4	2.17	0.63	0.11	0.00

Table 9. Mahalanobis D² Statistic for 8-17 July 1991, Total Cloud (Column and Row Labels are Group Numbers).

Grouping #1

	1	2	3	4
1	0.00	0.94	2.26	4.59
2	0.94	0.00	0.40	2.05
3	2.26	0.40	0.00	0.98
4	4.59	2.05	0.98	0.00

Grouping #2

	1	2	3	4
1	0.00	0.82	1.84	4.06
2	0.82	0.00	0.28	1.78
3	1.84	0.28	0.00	0.90
4	4.06	1.78	0.90	0.00

Grouping #3

	1	2	3	4
1	0.00	0.99	2.71	5.64
2	0.99	0.00	0.59	2.88
3	2.71	0.59	0.00	1.26
4	5.64	2.88	1.26	0.00

Grouping #4

	1	2	3	4
1	0.00	0.88	2.12	4.41
2	0.88	0.00	0.37	1.98
3	2.12	0.37	0.00	0.96
4	4.41	1.98	0.96	0.00

Table 10. 4-Category Cloud Amount Groupings.

Cloud Amount (%) Range for Category Index

Grouping #	1	2	3	4
1	0	5-40	45-80	85-100
2	0	5-30	35-65	70-100
3	0	5-45	50-95	100
4	0	5-35	40-75	80-100

various predictand categories. Also important is the degree to which the MDA assigns the highest probability to the correct category; that is, the category corresponding to the cloud amount that is associated with the independent values of the predictors at a gridpoint.

Miller[21] cites a measure of the accuracy of the MDA in assigning the correct category (of a designation of G categories) to a specified set of N predictor value vectors as

$$\overline{P} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{G} (P_{ij} - O_{ij})^2$$

where P_{ij} is the diagnosed probability at the gridpoint i in predictand category j, O_{ij} = 1 if category j occurred at gridpoint i, and O_{ij} = 0 if it did not. Thus, $\overline{P}$ is a sort of mean-squared accuracy of the diagnosed probability for a given set of independent predictor vectors and designated groupings of corresponding predictands.

In Table 11, we show the values of $\overline{P}$ for the 18, 21, and 24 July 1991 verifications given the four different groupings of the predictand. We can discern a relationship between the range (number of five-percent cloud amount categories) of the categories and the $\overline{P}$ values. In high cloud, which is dominated by 0 percent cloud amount, all groupings performed nearly equally because all four groupings had 0 percent in a category by itself. In middle and low cloud decks, $\overline{P}$ is lowest in Grouping 3 because the partly cloudy categories are most prevalent in middle and low clouds, and they have the greatest range in Grouping 3. The same holds true for total cloud which has greatest prevalence of overcast or near-overcast: because Grouping 2 had the widest range and Grouping 3 had the narrowest range of largest cloud amounts, they had the lowest and highest $\overline{P}$ values respectively. Thus, $\overline{P}$ may be a useful tool for assessing probability accuracy from case to case for a given grouping of the predictand, but it is not as useful for deciding which grouping might be best.

Table 11. $\overline{P}$ Values for 8-17 July 1991 for the 4 Category Groupings Shown in Tables 8 and 9.

Deck	Grouping 1	Grouping 2	Grouping 3	Grouping 4
H	0.494	0.494	0.490	0.496
M	0.676	0.682	0.638	0.683
L	0.686	0.682	0.645	0.689
T	0.579	0.536	0.613	0.566

Once probabilities for each predictand category are diagnosed at each gridpoint based on the vector of predictor values, it is necessary to select the most likely category of cloud amount. At gridpoints where one of the categories have a probability of nearly 1 (and thus the other category probabilities are nearly zero), it is a simple matter to select the most likely cloud amount category. If the MDA can clearly discriminate between the categories at all gridpoints, then selection of most likely category at all gridpoints is trivial: in all cases, it is the category with the highest probability. However, for most applications of MDA this unfortunately is not the case. In the worst case, probabilities for the categories can be nearly equal, leading to a virtually arbitrary decision as to which category to select. Thus it is necessary to devise and implement a category selection strategy based on the diagnosed probabilities for all the gridpoints.

In our project, we developed and tested five different category selection methods. The following discussion describes the five methods. Methods (3) - (5) are an attempt to preserve the frequency of occurrence of the observed cloud amounts in the diagnosed cloud amount distributions.

(1) Maximum probability method--at each gridpoint, the cloud amount category with the highest diagnosed probability is selected to represent the cloud amount range for that gridpoint.

(2) Weighted probability method--at each gridpoint, the mean value of the cloud amount range for each category is multiplied by the probability computed for the category, and the products are summed over the categories to give a probability-weighted cloud amount. The category in which the cloud amount falls is selected to represent the cloud amount range for that gridpoint.

(3) Frequency of occurrence sorted probability method--this procedure is carried out as follows: compute the average probability for each cloud amount category over all hemispheric gridpoints for a particular diagnosis time. Then order the categories from highest average probability to lowest. Next, compute the frequency of occurrence of observed cloud amounts over the 10-day development period for each category. Then for each category in order of their average probability: sort the gridpoint probabilities for that category, find the threshold probability such that the frequency of occurrence proportion of the hemispheric gridpoints have probabilities greater than the threshold, and assign that category value to each gridpoint having a category probability greater than the threshold. For successive categories, consider only gridpoints not already assigned to a category.

(4) Category selection using probability thresholds (Allen and Le Marshall[22])--a threshold probability is determined for each cloud amount category by sorting the probabilities from all hemispheric gridpoints for each category, then finding the probability threshold such that the frequency of occurrence proportion of the hemispheric gridpoints have a probability greater than the threshold. Then for each gridpoint, we start with the category having the lowest hemispheric average probability, and select the category in which the gridpoint's category probability first exceeds the category threshold. If one of the less likely categories is not selected, the most likely (highest average probability) category is selected.

(5) Iterative maximum probability method--the frequency of occurrence of observed cloud amounts over the 10-day development period is first computed for each cloud amount category. Next, the frequency of occurrence proportion of the number hemispheric gridpoints in each category is computed (for each category, NTOT = frequency of occurrence times the number of hemispheric points). Then the categories for each gridpoint are ordered from highest probability to lowest probability. On the first pass through the gridpoints, we identify the highest probability category for each gridpoint, compute the difference between that probability and the next highest probability, sort these probability differences over all gridpoints that have the same highest probability category, and select for that

74

category the NTOT gridpoints that have the largest differences. Once this is done for all categories, take a second pass through the unassigned gridpoints, and considering only those categories that are not full (that is, do not have NTOT gridpoints assigned to them), identify the gridpoints that have an unfilled category as their second highest probability. Within these unfilled categories, compute the difference between that probability and the next largest category's probability. Sort these differences, and select for that category up to NTOT gridpoints with the highest differences. Repeat this process until at least all but one category is filled, then assign any unassigned gridpoints to the remaining unfilled category.

We diagnosed probabilities for hemispheric gridpoints using 12-h forecast predictor values initialized on 18, 21, and 24 July 1991, at both 0 and 12 UTC. In each case, we used the 10-day development period just prior to each of the three respective dates to develop the discriminant functions. The probability diagnoses were conducted for all four 4-category groupings discussed earlier. Then for each 4-category grouping and each date, we applied all five category selection methods separately to obtain the cloud amount category values. Then for each 4-category grouping and category selection method pair, we verified the selected categories for all gridpoints for the three dates together against the same 4-category grouping of the observed cloud amounts at the forecast verification times.

Tables 12-15 show the category skill scores computed in the verifications of the MDA probability diagnosis and category selection processes. Each table compares the five category selection methods for a given 4-category grouping. The skill scores computed for these tables are: percentage of all verified points assigned to the correct category, category root-mean-square error (RMSE), and what we shall call the frequency of occurrence fit. The latter quantity is given by

$$\frac{1}{N} \sum_{g=1}^{G} \left| \left(m_g / n_g \right) - 1 \right| n_g$$

where g is the category index of G (here, four) categories, m_g is the number of gridpoints diagnosed to be in category g, n_g is the number of gridpoints observed in

that category (the verification points), and N is the total number of verification points. This is a measure of the agreement between the frequency of occurrence of the diagnosed cloud amounts for each category and the frequency of occurrence for the verification cloud amounts.

The category selection methods that do not make an attempt to preserve the observed frequency of occurrence of cloud amount, maximum probability, and weighted probability, clearly do not do so, as indicated by the frequency of occurrence fit. Because preserving frequency of occurrence of observed clouds was our primary motivation for considering MDA, we eliminated these two category selection methods from further consideration. In evaluating the other three category selection methods, we assigned one point for the best score in each deck/skill score entry separately for each of the groupings. By this standard, the iterative maximum probability method scored the best in Groupings 1, 2, and 4, and tied for the best in Grouping 3. In fact, the order of skill for the three methods was consistent among the four groupings. This suggests that the category selection method has more impact on skill in the MDA-type cloud diagnosis than does the choice of predictand groupings.

We next performed the cloud amount category diagnosis for the 12-hour forecast predictor values initialized on 18, 21, and 24 January 1991 at both 00 and 12 UTC. As usual, we used the 10-day development period just prior to each of the three dates to develop the discriminant functions. We performed discriminant function development (dependent data) and the cloud amount category probability diagnosis (independent data) using the 4-category Grouping 4. Next, all five category selection methods were applied to the resulting category probabilities to select the cloud amount category for each gridpoint. Then, the observed cloud amounts in the same 4-category grouping were used to perform the verification at the forecast valid times. The verification results are shown in Table 16, which is the January counterpart to Table 15.

Table 12. Comparison of Verification Scores of MDA Category Classification Methods, 12-Hour Forecasts Initialized on 18, 21, and 24 July 1991, Northern Hemisphere, 4-Category Grouping 1

Deck	Max. Prob.	Wghtd. Prob.	FOO Srtd. Prob.	Cat. Sel. Using Pr. Th.	Iter. Max. Prob.
Percent in the Correct Category (Perfect = 100)					
H	62.9	34.0	59.2	60.0	59.7
M	41.8	35.9	40.0	39.5	40.2
L	40.7	36.9	38.8	39.7	39.5
T	53.4	41.8	50.6	51.2	50.8
Category RMSE (Perfect = 0)					
H	0.991	0.930	0.942	0.960	0.923
M	1.037	0.968	1.138	1.122	1.132
L	1.063	0.947	1.152	1.123	1.141
T	1.069	0.924	1.033	1.057	1.027
Frequency of Occurrence Fit (Perfect = 0)					
H	0.400	1.033	0.027	0.140	0.024
M	0.304	0.835	0.097	0.261	0.079
L	0.246	0.757	0.069	0.184	0.088
T	0.261	0.694	0.104	0.190	0.088

Table 13. Comparison of Verification Scores of MDA Category Classification Methods, 12-Hour Forecasts Initialized on 18, 21, and 24 July 1991, Northern Hemisphere, 4-Category Grouping 2

Deck	Max. Prob.	Wghtd. Prob.	FOO Srtd. Prob.	Cat. Sel. Using Pr. Th.	Iter. Max. Prob.
Percent in the Correct Category (Perfect = 100)					
H	63.8	33.1	59.4	59.9	59.5
M	40.7	30.0	39.4	37.8	40.0
L	42.4	30.0	39.7	39.8	40.0
T	58.5	50.7	54.0	55.4	54.2
Category RMSE (Perfect = 0)					
H	1.101	0.987	1.008	1.045	1.006
M	1.295	1.046	1.197	1.283	1.196
L	1.240	1.017	1.212	1.219	1.188
T	1.129	0.961	1.065	1.078	1.061
Frequency of Occurrence Fit (Perfect = 0)					
H	0.457	1.021	0.027	0.154	0.023
M	0.309	1.022	0.038	0.091	0.083
L	0.395	1.064	0.079	0.106	0.090
T	0.387	0.436	0.086	0.202	0.075

Table 14. Comparison of Verification Scores of MDA Category Classification Methods, 12-Hour Forecasts Initialized on 18, 21, and 24 July 1991, Northern Hemisphere, 4-Category Grouping 3

Deck	Max. Prob.	Wghtd. Prob.	FOO Srtd. Prob.	Cat. Sel. Using Pr. Th.	Iter. Max. Prob.
Percent in the Correct Category (Perfect = 100)					
H	62.5	34.9	60.0	60.3	60.3
M	46.8	40.0	44.5	44.3	43.8
L	46.9	44.2	43.2	44.2	42.2
T	50.2	38.8	48.9	49.0	48.8
Category RMSE (Perfect = 0)					
H	0.884	0.874	0.862	0.864	0.837
M	0.914	0.880	0.988	0.971	0.996
L	0.879	0.860	1.004	0.972	1.020
T	0.925	0.883	0.995	0.917	0.975
Frequency of Occurrence Fit (Perfect = 0)					
H	0.341	1.070	0.033	0.104	0.023
M	0.209	0.671	0.097	0.134	0.077
L	0.346	0.524	0.070	0.110	0.054
T	0.123	0.813	0.098	0.103	0.079

Table 15. Comparison of Verification Scores of MDA Category Classification Methods, 12-Hour Forecasts Initialized on 18, 21, and 24 July 1991, Northern Hemisphere, 4-Category Grouping 4

Deck	Max. Prob.	Wghtd. Prob.	FOO Srtd. Prob.	Cat. Sel. Using Pr. Th.	Iter. Max. Prob.
Percent in the Correct Category (Perfect = 100)					
H	63.1	33.6	59.1	59.9	59.4
M	40.3	33.8	39.2	39.0	39.5
L	40.3	34.6	38.8	39.7	39.4
T	54.9	44.5	51.5	52.5	51.8
Category RMSE (Perfect = 0)					
H	1.035	0.949	0.958	0.996	0.952
M	1.107	0.994	1.170	1.158	1.152
L	1.104	0.967	1.169	1.123	1.155
T	1.087	0.929	1.056	1.066	1.039
Frequency of Occurrence Fit (Perfect = 0)					
H	0.434	1.024	0.027	0.148	0.023
M	0.234	0.892	0.103	0.073	0.081
L	0.131	0.890	0.078	0.153	0.107
T	0.314	0.625	0.104	0.191	0.083

Table 16. Comparison of Verification Scores of MDA Category Classification Methods, 12-Hour Forecasts Initialized on 18, 21, and 24 January 1991, Northern Hemisphere, 4-Category Grouping 4

Deck	Max. Prob.	Wghtd. Prob.	FOO Srtd. Prob.	Cat. Sel. Using Pr. Th.	Iter. Max. Prob.
			Percent in the Correct Category (Perfect = 100)		
H	70.3	37.8	64.6	65.6	64.9
M	42.6	31.8	39.8	40.3	40.7
L	42.9	32.1	41.3	41.1	42.3
T	49.9	35.6	47.9	47.1	48.2
			Category RMSE (Perfect = 0)		
H	0.935	0.894	0.876	0.877	0.849
M	1.179	0.988	1.216	1.236	1.171
L	1.110	0.979	1.172	1.237	1.155
T	1.163	0.980	1.112	1.182	1.110
			Frequency of Occurrence Fit (Perfect = 0)		
H	0.458	1.011	0.025	0.072	0.033
M	0.309	0.980	0.063	0.278	0.077
L	0.111	0.933	0.054	0.321	0.068
T	0.363	0.874	0.039	0.155	0.034

The January results show the same trends as the July results. The poor frequency of occurrence fit for the maximum and weighted probability methods eliminate them from consideration. Among the three remaining methods, Table 16 shows that the iterative maximum probability achieves the best scores in a majority of the deck/skill score entries. Also, in comparing the iterative maximum probability scores between the two months, we see that the scores are similar in magnitude.

The next step was to compare the cloud amount category diagnosis skill of the 4-category MDA with that of MLR as a reference. Up to this point, all verifications of MLR cloud amount diagnoses were done by comparing the diagnosed cloud amount, rounded to the nearest 5 percent, with the corresponding forecast valid time transformed RTNEPH cloud amount in each deck at each gridpoint. Because the MDA method operates on ranges or categories of the predictand, we must now perform the verification by placing each gridpoint value of both the MLR and

transformed RTNEPH cloud amounts in their appropriate categories, then perform the verification. It is not possible to convert the category MDA value into a specific percentage cloud amount, since the category represents a range of cloud amount values. One could designate the mean of the range as the cloud amount value to represent each category, but that is an arbitrary choice and puts the MDA diagnosis at a disadvantage in verification.

In Table 17, we present the cloud amount category verification scores for the three-day January and July 12-hour forecast cloud diagnoses for both MLR and MDA. We converted both MLR and transformed RTNEPH (at forecast times) cloud amounts into the Grouping 4 4-category designations, then performed the verifications of both MLR and MDA against the valid time transformed RTNEPH cloud amount category values. We added two new scores, bias and mean absolute error (MAE), to allow for a comparison of the methods for skill indicators similar to those used to evaluate MLR and REEP/MLR earlier in this report. Bias, RMSE, and MAE have direct counterparts in the earlier verifications. The "percent in correct category" score is analogous to "20/20" score in the cloud amount verifications, because the latter skill score indicates the fraction of points that have diagnosed cloud amounts that lie within a certain range (here, 20 percent) of the observed cloud amount. In "percent in correct category," we give the percentage of points that have a diagnosed cloud category that is within the range (the category range) of the observed cloud amount category. The frequency of occurrence fit score is analogous to, but more stringent than, the normalized sharpness score in the cloud amount verifications, because the latter skill score indicates the number of points that have diagnosed cloud amounts $\leq$ 20 percent or $\geq$ 80 percent divided by the number of points that have observed cloud amounts in these ranges. In frequency of occurrence fit, we compute the departure from unity of the ratio of diagnosed to observed points for each cloud amount category, and then weight each of these departures by the number of observed points in the respective category in computing the average departure. Thus, it is not just a measure of how well the

diagnosed cloud amount category frequency of occurrence matches that of the observed cloud amounts in the extreme categories, but in all categories.

The verification results in Table 17 reveal some fundamental features of both MLR- and MDA-diagnosed cloud amounts. First, the biases for both tend to be small, with values greater than 10 percent of a category index only for MLR in the

Table 17. Comparison of Category Skill Scores Between MLR and MDA 12-Hour Forecasts Initialized on 18, 21, and 24 January and July 1991, Northern Hemisphere, 4-Category Grouping 4.

Deck	January		July	
	MLR	MDA	MLR	MDA
Bias (Perfect = 0)				
H	0.248	-0.001	0.259	-0.024
M	0.090	0.083	0.130	-0.037
L	0.068	0.015	0.070	0.007
T	-0.050	0.042	-0.036	-0.053
RMSE (Perfect = 0)				
H	0.871	0.849	0.954	0.952
M	1.029	1.171	1.062	1.152
L	1.018	1.155	1.036	1.155
T	1.015	1.110	0.988	1.039
MAE (Perfect = 0)				
H	0.570	0.462	0.633	0.555
M	0.748	0.823	0.774	0.821
L	0.733	0.801	0.758	0.827
T	0.698	0.728	0.646	0.662
Percent in Correct Category (Perfect = 100)				
H	51.8	64.9	49.2	59.4
M	39.6	40.7	38.8	39.5
L	40.7	42.3	38.6	39.4
T	45.2	48.2	50.2	51.8
Frequency of Occurrence Fit (Perfect = 0)				
H	0.536	0.033	0.443	0.023
M	0.366	0.077	0.373	0.081
L	0.307	0.068	0.424	0.107
T	0.325	0.034	0.211	0.083

high (both months) and middle (July) cloud decks. The biases for MDA are smaller than those for MLR in all deck/month entries except one (total/July). In both RMSE and MAE, the MLR scores are better than the MDA scores in all but the high cloud deck for both months. This is consistent with earlier results that have shown the MLR consistently produces the lowest mean-squared errors of any diagnosis method. Unfortunately, as can be seen by the frequency of occurrence fit scores, this minimized mean-squared error is attained at the expense of the frequency of occurrence distribution of the diagnosed cloud amounts. The frequency of occurrence fit of MLR is from three to 19 times worse than that of MDA. As was seen in the earlier discussion of the MLR results, this is due to an overestimate of the number of points diagnosed to have cloud amounts near the middle of the 0-100 percent range, and an underestimation of the number of points having cloud amounts in the extremes of the range. Finally, we see that MDA scores better than MLR in "percent in correct category" in all deck/month entries.

A disadvantage of MDA is that the final diagnosed quantity is a category, or range, of cloud amounts rather than a single value. The diagnosed single value, as is available from MLR or REEP/MLR, cannot be considered as an exact amount of cloudiness to expect, but rather an indicator of which part of the spectrum of possible cloud amounts (0-100 percent) to expect. Nevertheless, if one could narrow the subrange indicated by the diagnosis somewhat from that indicated by 4 categories, the cloud amount category score may be more useful to the user. For example, instead of a cloud amount category value of 5-35 percent (second category of the 4-category grouping), it may be more useful to know if the cloud amount is more likely to be 5-20 percent or 25-40 percent, if this can be done without an appreciable decrease in the categorical probability.

For this reason, we tried dividing the predictand cloud amount into both 6-category and 8-category groupings, as follows:

6-category:	0	5-20	25-40	45-60	65-80	85-100		
8-category:	0	5-15	20-30	35-45	50-60	65-75	80-90	95-100

6-category:	0	5-20	25-40	45-60	65-80	85-100		
8-category:	0	5-15	20-30	35-45	50-60	65-75	80-90	95-100

Obviously, the probability of cloud amount category for a higher number of categories will be lower than that for a lower number of categories. This will result in loss of skill in selecting the correct category from the diagnosed probabilities (because there are more categories to select from). However, both MLR and MDA would decrease in skill in terms of category verification because there are now more possible categories, so the diagnosed category can be different from the observed category by a greater number of categories.

We sought to determine if we could realize a gain in skill <u>relative to MLR</u> by performing the MDA over a greater number of categories of the predictand. We used both the 6-category and 8-category groupings in the 10-day development period prior to days 18, 21, and 24 of January and July 1991 as before, and then diagnosed the category probabilities for 12-hour forecast predictors initialized at 00 and 12 UTC on these dates. The "iterative maximum probability" method was again used to select the cloud amount category from the categorical probabilities at each point. We then converted both MLR and transformed RTNEPH cloud amounts to the 6- and 8-category designations, and verified both MDA and MLR against transformed RTNEPH cloud amount categories as we did with the 4-category groupings.

The results of these verifications are given in Tables 18 and 19. As expected, the RMSE, MAE, and Percent in Correct Category scores declined steadily in skill with the increase in the number of categories. However, the decline in skill as seen in the bias and frequency of occurrence fit scores is small and is not consistent with the increase in number of categories.

The amount of decline in skill from four categories to six categories is given in Table 20, while the decline in skill between six categories and eight categories is given in Table 21. In all scores except RMSE, the values represent a simple difference between the respective table values, while in RMSE the values represent the square root of the difference between the squares of the RMSE values. We have

indicated the skill score/month entries where the MDA scores declined in skill less than MLR, indicating a gain in skill relative to MLR. As can be seen, 21 of the 40 MDA skill score/month entries improve relative to MLR in going from four categories to six categories in the cloud amount. In changing from six categories to eight categories in the cloud amount, the MDA skill score/month entries improve relative to MLR in only 13 of the 40 cases. The difference in skill decline between MDA and MLR is less than in the 6-category to 4-category comparison. This suggests that the peak in MDA cloud amount category diagnosis skill relative to MLR diagnosis skill (when represented in cloud amount categories) is at about six categories. For this reason, we chose a 6-category MDA to produce cloud amount category diagnoses for the full 7-day verification period in each month, and for 12-, 24-, 36-, and 48-hour forecast predictors.

7.2 MDA vs. MLR and REEP/MLR Statistical Results

In Tables 22 and 23 we see the scores for the cloud amount category verification of diagnosed cloud amounts from MLR, REEP/MLR-A, REEP/MLR-B, and of the MDA-diagnosed cloud amount category probability converted to cloud amount category. The difference between the MLR and MDA values of Table 22, 23, and those of Table 18 is that the former represents the full 7-day verification period. These sets of values have the same trends, indicating that the 3 days on which the category selection method and number of categories decisions were made were representative of the full verification period.

In comparing the category skill scores in Tables 22 and 23 for the four diagnosis methods, we see first that biases are relatively small for all methods. MDA has values greater than 0.1 for only one deck/month entry (middle/Jan) REEP/MLR-A has values greater than 0.1 for only two deck/month entries (total/Jan and high/Jul). MLR and REEP/MLR-B has more larger biases than the other two methods. In RMSE and MAE, MLR has the smallest errors in every deck/month entry except two (high deck in both months). Conversely, MDA has the largest

months). In "Percent in Correct Category" and "Frequency of Occurrence Fit," MDA has the best scores in virtually all deck/month entries, and MLR the worst.

We next broke down the "Percent in Correct Category" scores by observed cloud amount category. The results are shown in Figures 22 and 23. These figures indicate the percentage of gridpoints in each observed cloud amount category that were diagnosed correctly. It is clear from these figures that the MDA method excels in skill in the categories at the ends of the cloud amount range, while MLR performs best in the middle for the range. The two methods are more alike in skill in the transition categories. These observations shed more light on the relative capabilities of the methods than do the overall "Percent in Correct Category" scores in Tables 22 and 23.

We next broke down the "Frequency of Occurrence Fit" scores by category. We did this by graphing the frequency of occurrence of cloud amount categories for each method and for the (twice-daily) observed cloud amounts, for the 7-day verification period. These graphs are shown in Figures 24 and 25.

We compared the frequency of occurrence of the verification period twice-daily observed cloud amounts (on the graphs) with the entire month's four times/day observed cloud amounts (not shown) to see how representative the verification period frequency of occurrence values were. In both months, the 7-day values had somewhat lower frequency of occurrence in category one, and higher in category 6. The frequencies were more similar to each other in the other four categories. Thus the average cloud amount for the twice daily 7-day sample was higher than for the month as a whole. We did not determine if this was due to the seven days selected or the exclusion of the 06 and 18 UTC observed clouds in the sample. In any case, we noted that the trends of the frequency of occurrence from category to category were the same for both the 7-day sample and the entire month in each cloud deck.

In examining Figures 24 and 25, we see that MLR produces too few gridpoints having category one or six cloud amount values and too many gridpoints having category three or four values, and a mix of too many/too few gridpoints having two or five values. At the other extreme, MDA produces more nearly the correct

Table 18. Comparison of Category Skill Scores Between MLR and MDA 12-Hour Forecasts
Initialized on 18, 21, and 24 January and July 1991, Northern Hemisphere, 6-Category Grouping

Deck	January		July	
	MLR	MDA	MLR	MDA
Bias (Perfect = 0)				
H	0.310	-0.013	0.366	-0.038
M	0.088	0.144	0.187	-0.080
L	0.080	0.038	0.115	-0.004
T	-0.098	0.083	-0.047	-0.071
RMSE (Perfect = 0)				
H	1.334	1.363	1.514	1.541
M	1.710	1.902	1.786	1.977
L	1.706	1.909	1.771	1.975
T	1.716	1.874	1.683	1.775
MAE (Perfect = 0)				
H	0.854	0.722	0.993	0.881
M	1.277	1.355	1.351	1.441
L	1.271	1.352	1.360	1.460
T	1.216	1.253	1.127	1.159
Percent in Correct Category (Perfect = 100)				
H	46.9	61.8	43.4	55.5
M	28.4	32.6	26.4	29.7
L	29.1	38.8	24.5	28.6
T	34.8	40.3	41.3	43.7
Frequency of Occurrence Fit (Perfect = 0)				
H	0.535	0.040	0.446	0.033
M	0.416	0.075	0.440	0.078
L	0.358	0.071	0.541	0.075
T	0.337	0.083	0.264	0.086

Table 19. Comparison of Category Skill Scores Between MLR and MDA 12-Hour Forecasts Initialized on 18, 21, and 24 January and July 1991, Northern Hemisphere, 8-Category Grouping

Deck	January		July	
	MLR	MDA	MLR	MDA
Bias (Perfect = 0)				
H	0.355	-0.020	0.449	-0.052
M	0.110	0.190	0.212	-0.125
L	0.096	0.059	0.110	0.020
T	-0.174	0.109	-0.096	-0.086
RMSE (Perfect = 0)				
H	1.724	1.794	1.988	2.042
M	2.274	2.534	2.350	2.612
L	2.278	2.545	2.355	2.652
T	2.339	2.560	2.306	2.475
MAE (perfect = 0)				
H	1.087	0.940	1.298	1.159
M	1.712	1.810	1.792	1.914
L	1.722	1.833	1.832	1.993
T	1.700	1.759	1.595	1.670
Percent in Correct Category (Perfect = 100)				
H	44.8	60.2	40.5	53.2
M	23.4	28.2	21.9	25.5
L	23.7	28.3	19.5	22.9
T	27.7	32.9	34.0	36.2
Frequency of Occurrence Fit (Perfect = 0)				
H	0.536	0.036	0.443	0.037
M	0.416	0.088	0.418	0.111
L	0.367	0.071	0.493	0.083
T	0.353	0.047	0.264	0.099

Table 20. Decline in Category Skill Scores From 4-Category to 6-Category 12-Hour Forecasts
Initialized on 18, 21 and 24 January and July 1991, Northern Hemisphere

	January		July	
Deck	MLR	MDA	MLR	MDA
		Bias (Perfect = 0)		
H	0.062	0.012*	0.107	0.014*
M	-0.002	0.061	0.057	0.043*
L	0.012	0.023	0.045	-0.003*
T	0.048	0.041*	0.011	-0.018
		RMSE (Perfect = 0)		
H	1.010	1.066	1.176	1.212
M	1.366	1.499	1.436	1.607
L	1.369	1.520	1.436	1.602
T	1.384	1.510	1.362	1.439
		MAE (Perfect = 0)		
H	0.284	0.260*	0.360	0.326*
M	0.683	0.532*	0.577	0.620
L	0.538	0.551	0.602	0.633
T	0.518	0.525	0.481	0.497
		Percent in Correct Category (Perfect = 100)		
H	4.9	3.1*	5.8	3.9*
M	11.2	8.1*	12.4	9.8*
L	11.6	8.5*	14.1	10.8*
T	10.4	7.9*	8.9	8.1*
		Frequency of Occurrence Fit (Perfect = 0)		
H	-0.001	0.007	0.003	0.010
M	0.050	-0.002*	0.067	-0.003*
L	0.051	0.003*	0.117	-0.032*
T	0.012	0.049	0.053	0.003*

*MDA declined in skill less than MLR (21 score/month entries)

Table 21. Decline in Category Skill Scores From 6-Category to 8-Category 12-Hour Forecasts Initialized on 18, 21, and 24 January and July 1991, Northern Hemisphere

Deck	January		July	
	MLR	MDA	MLR	MDA
Bias (Perfect = 0)				
H	0.045	0.007*	0.083	0.014*
M	0.022	0.046	0.025	0.045
L	0.016	0.021	-0.005	0.016
T	0.076	0.026*	0.049	0.015*
RMSE (Perfect = 0)				
H	1.092	1.166	1.288	1.340
M	1.499	1.674	1.527	1.707
L	1.510	1.683	1.552	1.770
T	1.589	1.744	1.576	1.725
MAE (Perfect = 0)				
H	0.233	0.218*	0.305	0.278*
M	0.435	0.455	0.441	0.473
L	0.451	0.481	0.472	0.533
T	0.484	0.506	0.468	0.511
Percent in Correct Category (Perfect = 100)				
H	2.1	1.6*	2.9	2.3*
M	5.0	4.4*	4.5	4.2*
L	5.4	5.5	5.0	5.7
T	7.1	7.4	7.3	7.5
Frequency of Occurrence Fit (Perfect = 0)				
H	0.001	-0.004*	-0.003	0.004
M	0.000	0.013	-0.022	0.033
L	0.009	0.000*	-0.048	0.008
T	0.016	-0.036*	0.000	0.013

*MDA declined in skill less than MLR (13 score/month entries)

Table 22. Comparison of Category Skill Scores Among Diagnosis Methods, 12-Hour Forecasts for the Verification Period 18-24 January 1991, Northern Hemisphere, 6-Category Grouping

Deck	MLR	REEP/MLR-A	REEP/MLR-B	MDA
		Bias (Perfect = 0)		
H	0.273	0.010	-0.130	-0.025
M	0.090	-0.098	-0.180	0.177
L	0.092	-0.079	-0.201	0.062
T	-0.133	-0.148	-0.203	0.058
		RMSE (Perfect = 0)		
H	1.354	1.379	1.344	1.402
M	1.703	1.816	1.798	1.893
L	1.702	1.834	1.787	1.907
T	1.719	1.786	1.946	1.866
		MAE (Perfect = 0)		
H	0.863	0.759	0.714	0.743
M	1.268	1.315	1.287	1.345
L	1.264	1.330	1.282	1.342
T	1.217	1.242	1.343	1.242
		Percent in Correct Category (Perfect = 100)		
H	46.9	60.1	62.3	61.5
M	28.6	31.0	32.2	33.0
L	29.5	30.9	32.5	34.5
T	34.8	36.6	36.3	40.7
		Frequency of Occurrence Fit (Perfect = 0)		
H	0.537	0.147	0.114	0.041
M	0.419	0.227	0.219	0.081
L	0.378	0.281	0.287	0.061
T	0.359	0.250	0.172	0.041

Table. 23. Comparison of Category Skill Scores Among Diagnosis Methods, 12-Hour Forecasts for the Verification Period 18-24 July 1991, Northern Hemisphere, 6-Category Grouping

Deck	MLR	REEP/MLR-A	REEP/MLR-B	MDA
		Bias (Perfect = 0)		
H	0.345	-0.141	-0.098	-0.079
M	0.213	0.020	0.056	-0.047
L	0.158	0.023	0.099	0.064
T	-0.048	-0.083	0.098	-0.068
		RMSE (Perfect = 0)		
H	1.521	1.656	1.497	1.559
M	1.796	1.935	1.948	1.982
L	1.774	1.883	1.912	1.973
T	1.681	1.721	1.828	1.767
		MAE (Perfect = 0)		
H	0.997	0.975	0.872	0.897
M	1.358	1.434	1.452	1.446
L	1.360	1.421	1.450	1.459
T	1.122	1.144	1.207	1.150
		Percent in Correct Category (Perfect = 100)		
H	43.4	52.9	54.9	55.0
M	26.2	27.4	27.1	29.5
L	24.6	25.4	24.8	28.3
T	41.5	41.6	42.0	43.9
		Frequency of Occurrence Fit (Perfect = 0)		
H	0.431	0.209	0.166	0.052
M	0.426	0.282	0.342	0.068
L	0.520	0.390	0.368	0.068
T	0.255	0.255	0.131	0.086

Table 24. Cloud Amount Category Skill Scores for the Verification Period 18-24 January and July 1991, Northern Hemisphere

		January				July			
Deck	Frcst Hr	MLR	RMLR-A	RMLR-B	MDA	MLR	RMLR-A	RMLR-B	MDA
					Bias				
H	12	0.273	0.010	-0.130	-0.025	0.345	-0.141	-0.098	-0.079
	24	0.288	-0.256	-0.133	-0.012	0.340	0.031	-0.116	-0.079
	36	0.277	-0.261	-0.139	0.004	0.339	-0.004	-0.118	-0.048
	48	0.248	-0.272	-0.202	-0.002	0.332	0.005	-0.123	-0.037
M	12	0.090	-0.098	-0.180	0.177	0.213	0.002	0.056	-0.047
	24	0.101	-0.044	-0.205	0.158	0.201	0.007	0.074	-0.003
	36	0.085	-0.078	-0.215	0.176	0.203	0.011	0.045	-0.060
	48	0.074	-0.134	-0.214	0.174	0.197	0.016	0.058	-0.020
L	12	0.092	-0.079	-0.201	0.062	0.158	0.023	0.099	0.064
	24	0.091	-0.096	-0.157	0.098	0.133	0.018	0.201	0.103
	36	0.108	-0.087	-0.166	0.123	0.149	-0.018	0.141	0.028
	48	0.090	-0.095	-0.156	0.126	0.147	0.020	0.175	0.109
T	12	-0.133	-0.148	-0.203	0.058	-0.048	0.083	0.098	-0.068
	24	-0.116	-0.148	-0.139	0.077	-0.074	-0.093	0.087	0.000
	36	-0.083	-0.128	-0.105	0.113	-0.025	-0.063	0.026	-0.027
	48	-0.109	-0.145	-0.076	0.105	-0.011	-0.049	0.093	0.003
					RMSE				
H	12	1.354	1.379	1.344	1.402	1.521	1.656	1.497	1.559
	24	1.372	1.544	1.376	1.436	1.543	1.597	1.528	1.608
	36	1.372	1.556	1.393	1.458	1.560	1.634	1.570	1.650
	48	1.377	1.575	1.376	1.488	1.584	1.654	1.600	1.693
M	12	1.703	1.816	1.798	1.893	1.796	1.935	1.948	1.982
	24	1.714	1.791	1.819	1.906	1.796	1.929	1.977	1.998
	36	1.722	1.824	1.805	1.936	1.818	1.952	1.989	2.036
	48	1.725	1.858	1.822	1.927	1.817	1.957	2.003	2.050
L	12	1.702	1.834	1.787	1.907	1.774	1.883	1.912	1.973
	24	1.720	1.846	1.838	1.936	1.781	1.886	1.938	1.986
	36	1.737	1.865	1.835	1.960	1.781	1.896	1.934	1.995
	48	1.732	1.878	1.848	1.958	1.791	1.891	1.953	2.000
T	12	1.719	1.786	1.946	1.866	1.681	1.721	1.828	1.767
	24	1.721	1.815	1.897	1.887	1.687	1.722	1.867	1.782
	36	1.757	1.843	1.932	1.923	1.703	1.744	1.806	1.807
	48	1.766	1.860	1.907	1.934	1.728	1.777	1.927	1.837

Table 24. Cloud Amount Category Skill Scores for the Verification Period 18-24 January and July 1991, Northern Hemisphere

		January				July			
Deck	Frcst Hr	MLR	RMLR-A	RMLR-B	MDA	MLR	RMLR-A	RMLR-B	MDA
					MAE				
H	12	0.863	0.759	0.714	0.743	0.997	0.975	0.872	0.897
	24	0.891	0.846	0.737	0.765	1.025	0.954	0.897	0.932
	36	0.903	0.865	0.746	0.782	1.044	0.985	0.924	0.961
	48	0.919	0.882	0.738	0.803	1.075	1.013	0.953	0.996
M	12	1.268	1.315	1.287	1.345	1.358	1.434	1.452	1.446
	24	1.286	1.300	1.306	1.361	1.373	1.433	1.480	1.468
	36	1.289	1.316	1.295	1.377	1.393	1.460	1.494	1.502
	48	1.297	1.345	1.311	1.379	1.405	1.466	1.515	1.516
L	12	1.264	1.330	1.282	1.342	1.360	1.421	1.450	1.459
	24	1.286	1.338	1.326	1.369	1.368	1.424	1.455	1.469
	36	1.303	1.365	1.327	1.390	1.370	1.437	1.468	1.481
	48	1.308	1.383	1.341	1.391	1.380	1.437	1.472	1.486
T	12	1.217	1.242	1.343	1.242	1.122	1.144	1.207	1.150
	24	1.229	1.273	1.311	1.260	1.141	1.156	1.245	1.159
	36	1.257	1.292	1.337	1.291	1.160	1.178	1.204	1.184
	48	1.275	1.310	1.329	1.304	1.188	1.211	1.305	1.209
					Percent in Correct Category				
H	12	46.9	60.1	62.3	61.5	43.4	52.9	54.9	55.0
	24	45.1	57.5	61.6	61.2	41.9	53.0	54.1	54.3
	36	43.7	56.0	61.6	60.7	40.9	51.9	53.7	53.8
	48	42.1	55.2	61.5	60.1	39.2	50.7	52.7	52.7
M	12	28.6	31.0	32.2	33.0	26.2	27.4	27.1	29.5
	24	27.7	31.2	31.6	32.3	25.3	27.3	26.5	28.7
	36	27.8	31.4	32.1	32.6	24.7	26.4	26.0	28.2
	48	27.4	30.6	31.6	31.9	24.0	26.5	25.4	27.8
L	12	29.5	30.9	32.5	34.5	24.6	25.4	24.8	28.3
	24	28.4	31.0	31.7	34.0	24.6	25.4	25.5	28.5
	36	28.2	29.9	31.5	33.8	24.3	24.8	24.6	27.7
	48	27.5	29.0	31.3	33.6	24.1	24.9	25.1	27.9
T	12	34.8	36.6	36.3	40.7	41.5	41.6	42.0	43.9
	24	33.9	35.5	36.3	40.5	40.2	40.6	40.7	44.0
	36	33.6	35.4	36.1	40.2	39.3	39.8	40.9	43.1
	48	32.4	34.9	35.6	39.6	38.3	38.9	39.0	42.8

Table 24. Cloud Amount Category Skill Scores for the Verification Period 18-24 January and July 1991, Northern Hemisphere

		January				July			
Deck	Frcst Hr	MLR	RMLR-A	RMLR-B	MDA	MLR	RMLR-A	RMLR-B	MDA
		Frequency of Occurrence Fit							
H	12	0.537	0.147	0.114	0.041	0.431	0.209	0.166	0.052
	24	0.590	0.145	0.118	0.036	0.473	0.232	0.165	0.063
	36	0.629	0.164	0.122	0.031	0.506	0.217	0.159	0.042
	48	0.679	0.180	0.119	0.035	0.531	0.218	0.160	0.052
M	12	0.419	0.227	0.219	0.081	0.426	0.282	0.342	0.068
	24	0.443	0.262	0.219	0.088	0.475	0.294	0.352	0.099
	36	0.434	0.226	0.221	0.088	0.484	0.305	0.366	0.079
	48	0.454	0.233	0.227	0.098	0.534	0.317	0.405	0.090
L	12	0.378	0.281	0.287	0.061	0.520	0.390	0.368	0.068
	24	0.389	0.317	0.298	0.080	0.507	0.364	0.291	0.059
	36	0.397	0.298	0.328	0.080	0.531	0.397	0.371	0.068
	48	0.426	0.325	0.322	0.082	0.533	0.392	0.295	0.074
T	12	0.359	0.250	0.172	0.041	0.255	0.255	0.131	0.086
	24	0.387	0.273	0.176	0.046	0.290	0.272	0.135	0.085
	36	0.380	0.266	0.160	0.059	0.322	0.292	0.170	0.099
	48	0.419	0.268	0.210	0.051	0.346	0.298	0.122	0.091
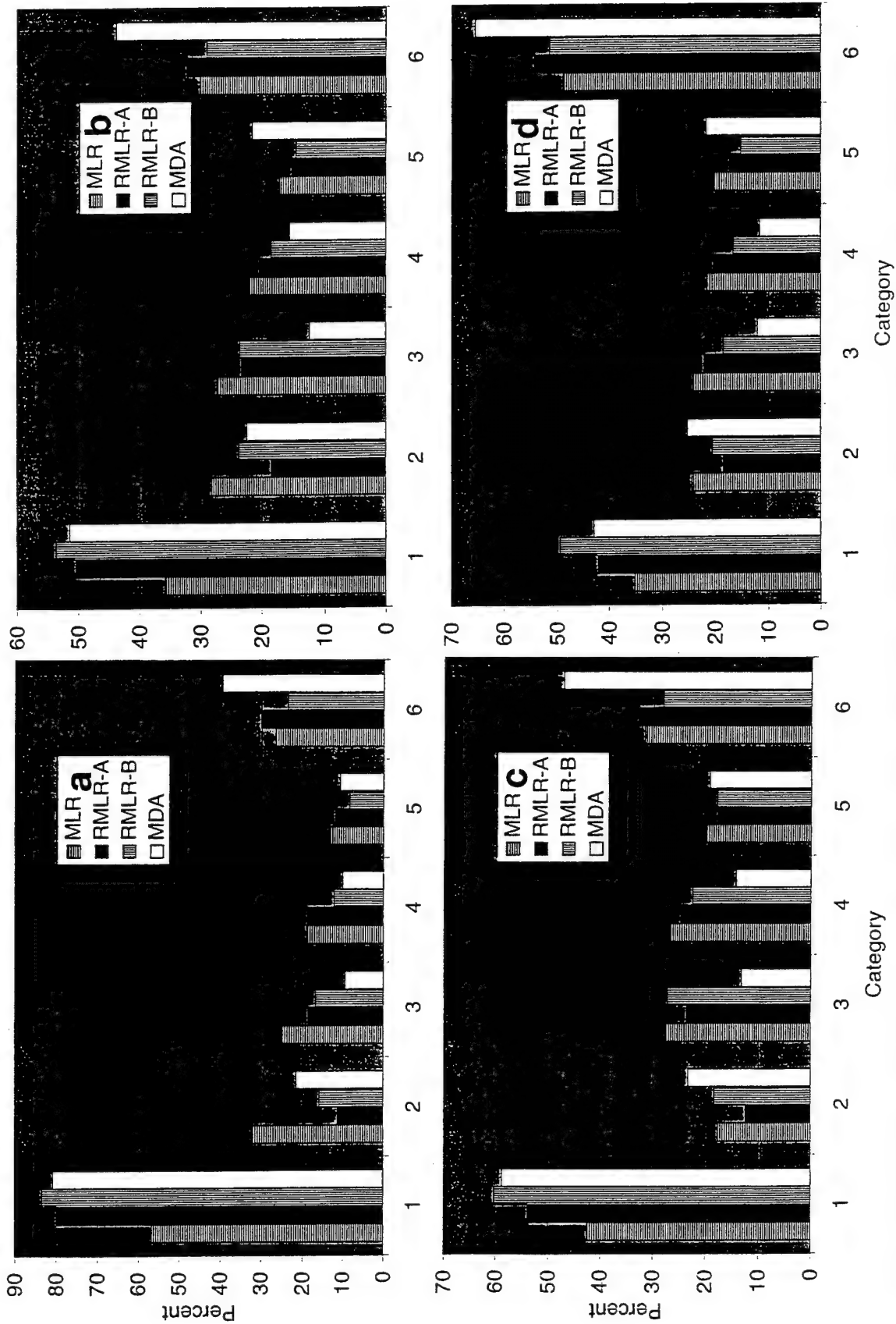
Figure 22. Percent Diagnosed in the Correct Cloud Amount Category Displayed by Observed Cloud Amount Category, 12-Hour Forecasts for the Verification Period 18-24 January 1991, Northern Hemisphere, for (a) High, (b) Middle, (c) Low, and (d) Total Deck Clouds.

95
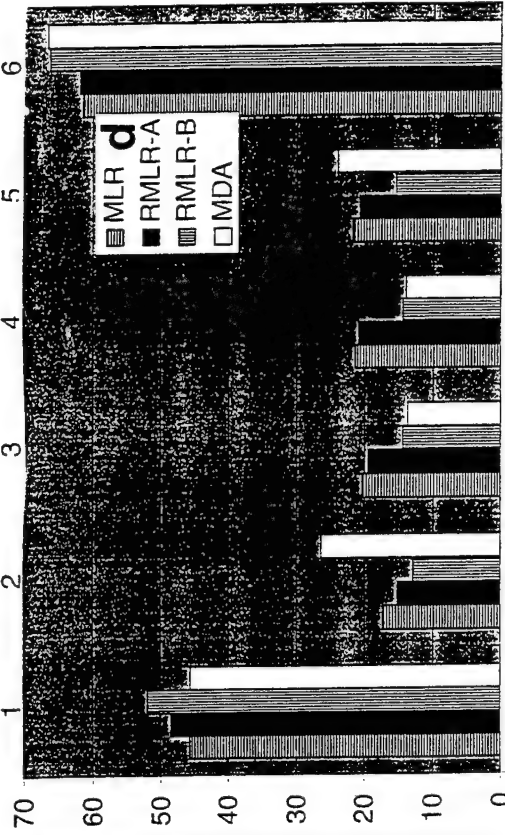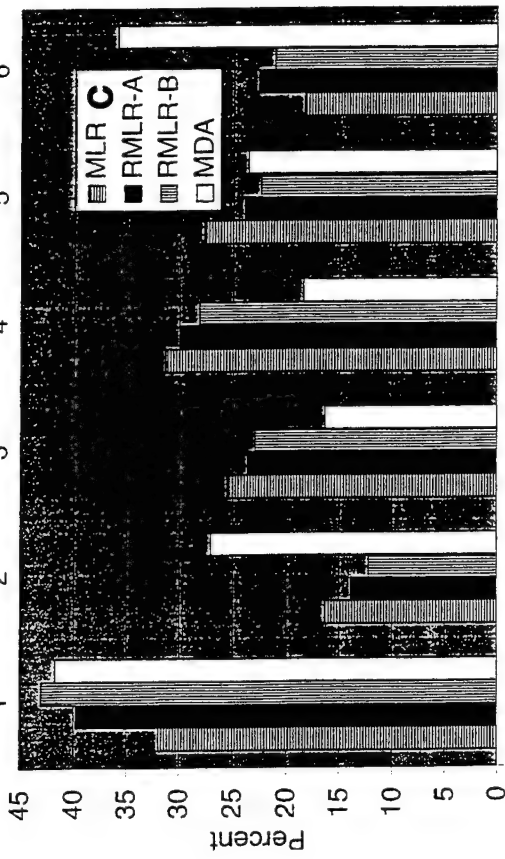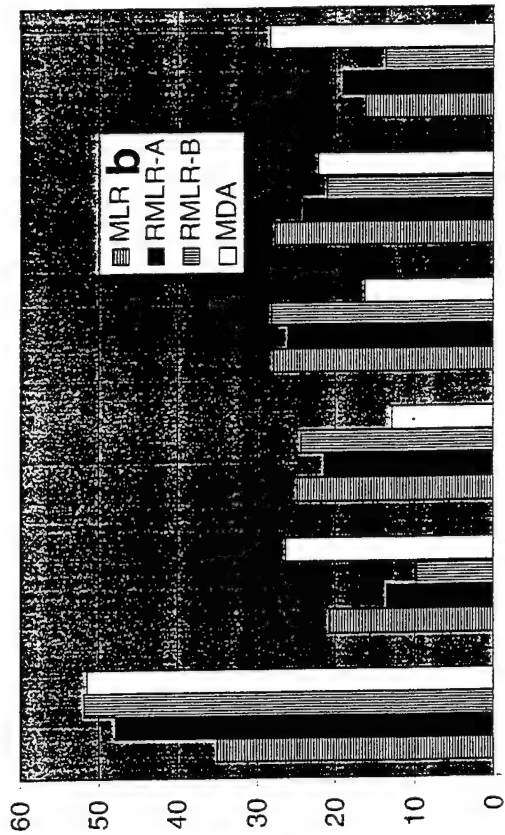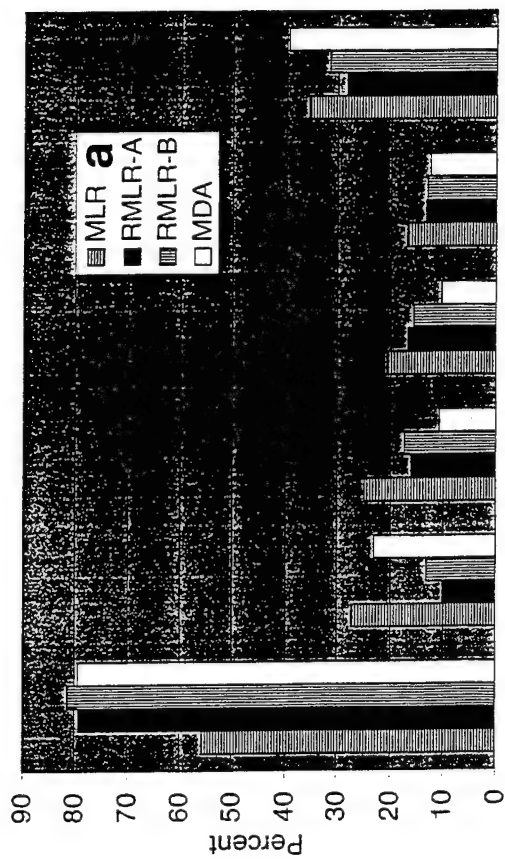
Figure 23. Same as in Figure 22 for Verification Period 18-24 July 1991.

96

number of gridpoints in all categories in the cloud amount category diagnosis. The REEP/MLR-A, B methods force enough category one gridpoints to be diagnosed, but in doing so reduce the number of category two gridpoints to below what is diagnosed by MLR. The over-estimate of categories three and four by MLR is reduced by the REEP/MLR methods, but not enough to coincide with the observed frequency of occurrence as well as MDA. The REEP/MLR methods boost the number of category 6 gridpoints diagnosed from that of MLR in all but the high deck, but only modestly.

The reason that MDA gets a higher total (over all categories) percent-in-correct-category score (see Tables 22 and 23) is that it outperforms the other diagnosis methods in this score in categories one and six. We now see that the observed cloud amount categories have their highest frequency of occurrence in these two categories. Since MDA has the best accuracy in the categories most likely to occur, MDA has the overall advantage.

The "iterative maximum probability" category selection method used in conjunction with MDA was designed to replicate the frequency of occurrence of observed clouds during the 10-day development period. This causes the resulting MDA-diagnosed frequencies to be more like those of observed clouds than the MLR or REEP/MLR-A, B.

As previously mentioned, an additional benefit of the MDA method is that it produces a probability for each cloud amount category. This probability is an estimate of the likelihood of each category being the correct one. The probability of the category that was selected for each gridpoint was evaluated as to the extra information it may contribute to the cloud amount estimate. We divided the probability range of 0-100 percent into six categories (0-4, 5-24, 25-44, 45-64, 65-84, 85-100) and assigned the probability of the cloud amount category selected at each gridpoint in the 7-day verification period MDA diagnoses to these probability categories. We first computed the frequency of occurrence of each probability category during the January and July verification periods. We next computed the percent of gridpoints in each probability category that had correctly diagnosed cloud
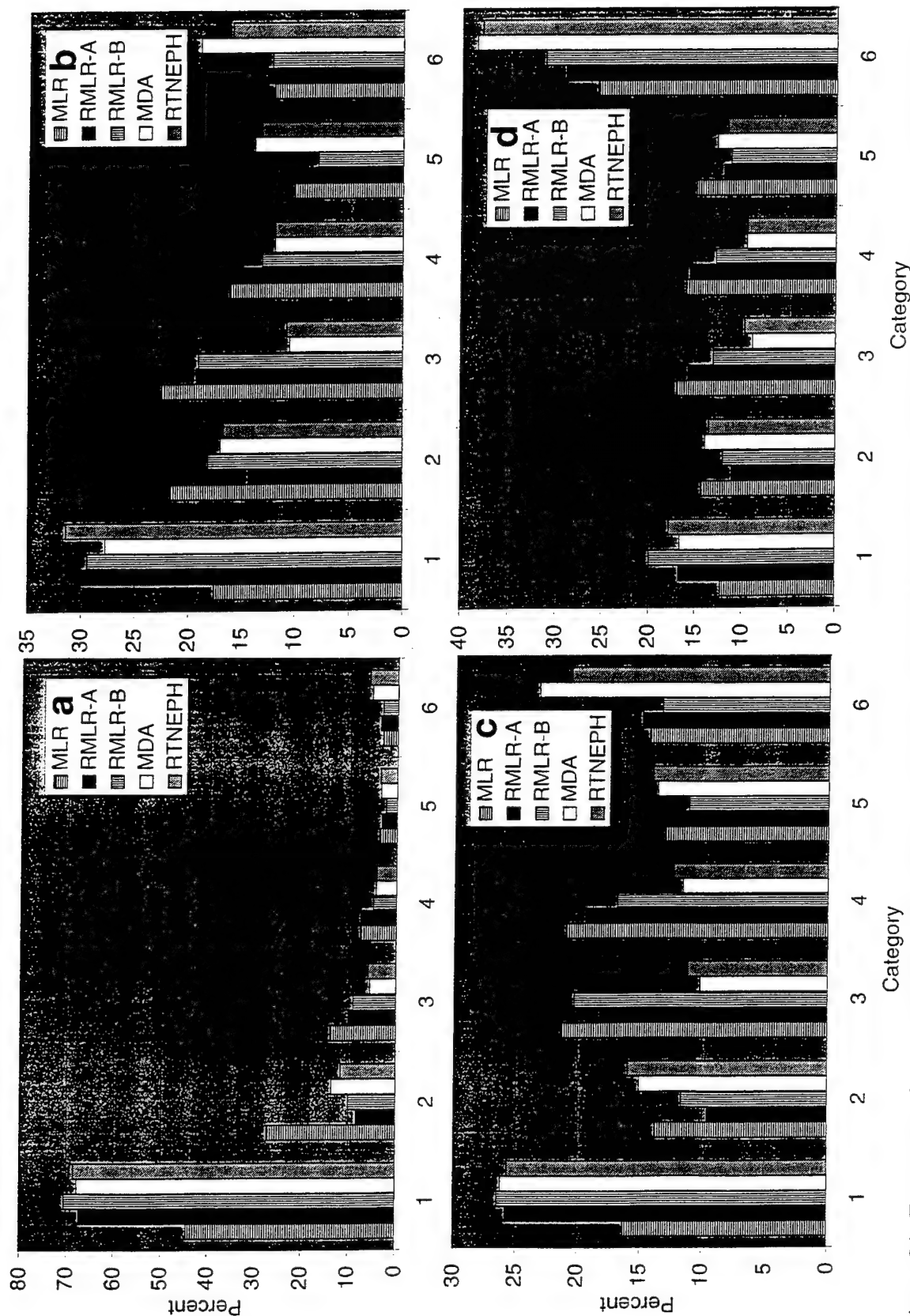
Figure 24. Frequency of Occurrence of Cloud Amount Categories, 12-Hour Forecasts and Transformed RTNEPH for the Verification Period 18-24 January 1991, Northern Hemisphere, for (a) High, (b) Middle, (c) Low, (d) Total Deck Clouds.

98
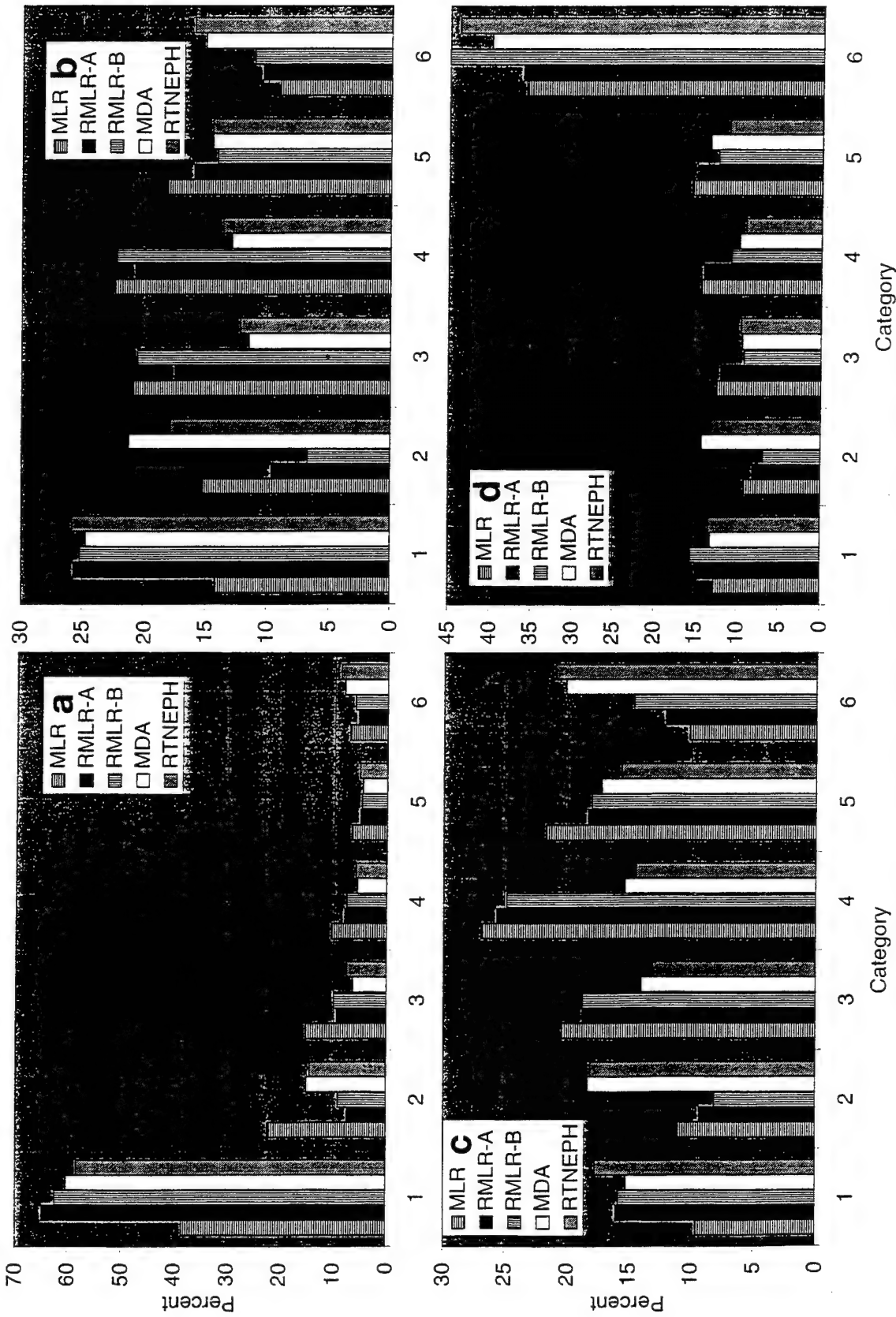
Figure 25. Same as in Figure 24 for Verification Period 18-24 July 1991.

amount categories by the MDA method. Both of these quantities are graphed for both months in Figure 26.

With the exception of the high cloud deck, the frequency of occurrence of probability category 2 (5-24 percent) is the highest in all cases. As expected, the frequency of occurrence decreases with increasing probability category index. In high cloud, probability categories 5 and 6 are frequent because 0 percent cloud amount is so likely [see Figure 24(a), RTNEPH]. In all of the decks, the percent of gridpoints with correctly diagnosed cloud amount categories increases with probability category index. This suggests that though there are fewer gridpoints diagnosed with high probabilities, those that are are very likely to be correctly diagnosed. Thus, when we see high probabilities associated with MDA-diagnosed cloud amount categories, we can feel confident that the predictions at those gridpoints are likely to be correct.

We next look at the cloud amount category skill scores from the 7-day verification periods as a function of forecast duration. In Table 24 we show these scores for the four diagnosis methods: MLR, REEP/MLR-A, REEP/MLR-B, and MDA. Note that the 12-hour forecast values are the same as those shown in Tables 22 and 23. We see the same relative abilities of the methods seen at 12-hours maintained at all later forecast durations. In bias and frequency of occurrence fit, there is no evident decrease in skill with forecast duration. In RMSE, MAE, and percent in correct category, there is a slow but steady decline in skill with increasing forecast duration in all diagnosis methods. MDA declined an average of 3 percent (of the 12-hour value) from 12- to 48-hours in percent in correct category. MLR declined an average of 7 percent in this skill score.

7.3 MDA vs. MLR: Cloud Amount Map Comparisons

As we did when evaluating the REEP/MLR methods against MLR, we must look at case examples of the performance of MDA with respect to MLR forecasts. This

Figure 26. MDA Frequency of Occurrence of Probabilities and Percent of Gridpoints With Correctly Diagnosed Cloud Amount Categories, Displayed by Diagnosed Probability Category, 12-Hour Forecasts for the Verification Periods 18-24 January and July 1991, Northern Hemisphere, for (a) High, (b) Middle, (c) Low, and (d) Total Deck Clouds.
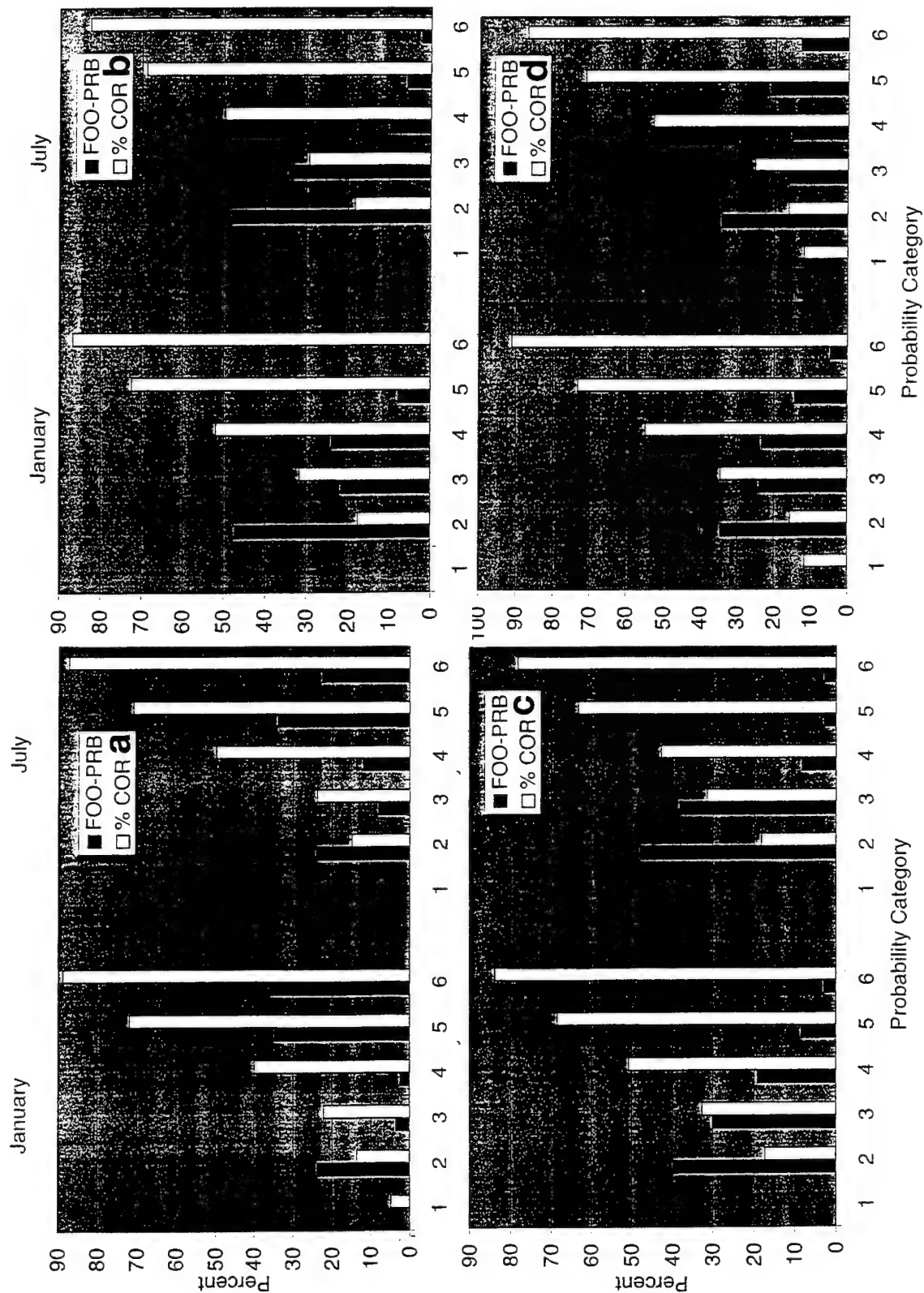
101

requires that map comparisons of the diagnosed cloud amounts from both methods be compared with the transformed RTNEPH amount distributions as a verification.

Since MDA produces predictions of the cloud amount category rather than the actual cloud amount, we must either transform the category value to a cloudamount or express the RTNEPH and MLR cloud amounts as category values. We did the latter in the statistical evaluations just discussed. However, to maintain continuity of the type of map comparisons shown, we converted MDA cloud categories into corresponding cloud amounts. For our purposes, we set the cloud amount equal to the average value in the category range of each cloud category, rounded to the nearest 5 percent. In the 6-category case, this resulted in cloud amount values of 0, 15, 35, 55, 75, and 95 percent. While this puts MDA-predicted cloud amounts at a disadvantage in a statistical comparison, it has little impact in map displays due to the qualitative nature of the maps.

In the following discussions, we will focus on the comparison of MLR and MDA cloud maps. In each case, we have included a map of the percent probability of the MDA-diagnosed cloud amount.

7.3.1 0000 UTC 25 JANUARY 1991 EUROPE CASE

Figure 27 shows the high cloud amount distribution over Europe on the subject day as depicted by MLR and MDA 12-hour forecasts, and by the transformed RTNEPH. The map also shows the percent probability obtained from the MDA method for the cloud amount category selected. Note this is not the probability of cloudiness–it is the probability of the cloud amount category selected being the correct one, as computed by the MDA method. We see that the diagnosis produced by MDA is more similar to the MLR forecast than to the RTNEPH. The major differences between the MLR and MDR diagnoses are the absence of cloudiness over northwestern Russia in the latter, and the smaller area of cloudiness over and northeast of the Black Sea. Also, it appears that the blob of cloudiness in the western Mediterranean is smaller in the MDA depiction. All three of these

102

differences favor MDA over MLR in comparison with RTNEPH. Notice in the probability map that the zones of low probability of cloud amount category coincide with the areas of diagnosed high cloud in MDA. This is understandable since a 0 percent cloud amount category is so prevalent in the high cloud analyses and is thus the most likely event.

The middle deck clouds from MLR and MDA (Figure 28) also are more similar to each other than to the transformed RTNEPH. The major areas of cloudiness in the western Mediterranean and in eastern Europe follow very similar spatial patterns, but MDA appears to have slightly larger areas of > 80 percent cloudiness than does MLR. The size of the < 20 percent cloud amount area over western and central Europe are about the same size and shape. These two facts result in sharper gradients in the MDA diagnosis. This is also true to a lesser extent over Scandinavia. The locations of the cloudy areas are not better placed in MDA--in fact, the larger areas of > 80 percent cloudiness in the southwestern and southeastern portions of the region are more exaggerated compared to RTNEPH than are those of MLR. In the probability map, we see that the only areas that have > 50 percent probability (of correct cloud amount category) are within the areas in which all three depictions are virtually cloud free. This suggests a usefulness to the probability distribution map–in the absence of the verification, the agreement between MLR and MDA in areas of highest probability suggests that these areas are the best bet for the highest accuracy of the predicted cloud amounts. The tighter gradients produced in the MDA diagnoses are even more apparent in the low cloud distribution maps (Figure 29). Over the North Sea, the appearance of the > 80 percent area in the MDA represents an improvement over MLR. Over the Black Sea and to its north and northeast, the MDA's larger > 80 percent area represents a decline in skill from MLR. This is also true in the western Mediterranean. Neither method captures the cloudiness in central Europe shown in the RTNEPH depiction. Notice in the probability map that almost no areas are characterized by > 50 percent probability. This may result from the fact that there are almost no significant areas where all three cloud depictions agree substantially.
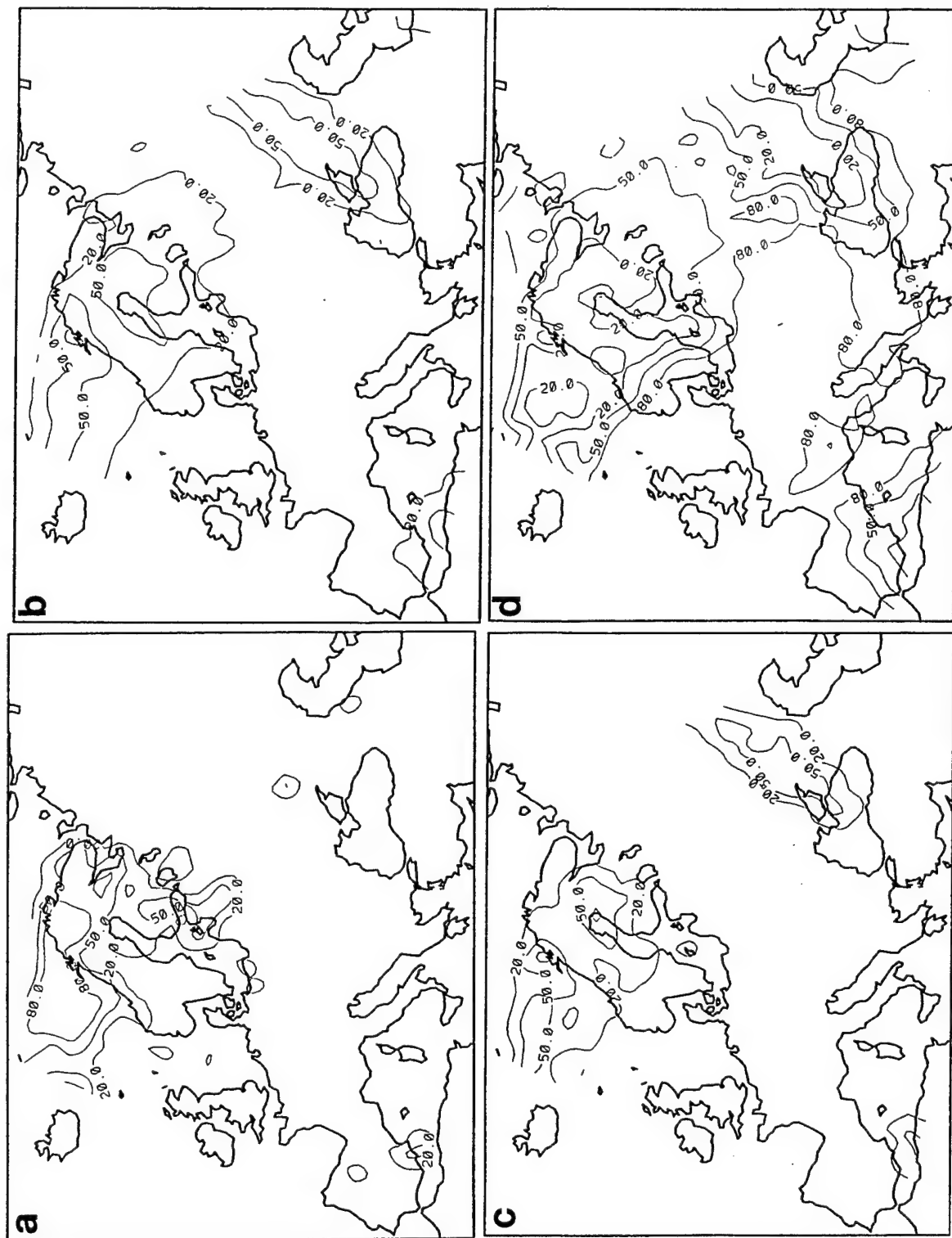
Figure 27. High Deck Cloud Amounts (%) Over Europe on 0000 UTC 25 January 1991 from (a) Transformed RTNEPH, (b) 12-Hour MLR Forecast, (c) 12-Hour MDA Forecast and (d) 12-Hour MDA Forecast Probabilities.

104

Figure 28. Same as in Figure 27 for Middle Deck Cloud Amounts.

105

The MDA total cloud depiction in Figure 30 presents a very cloudy scene, with very few areas of < 20 percent cloud amount. This contrasts with the MLR depiction which has few but somewhat larger areas of 20 percent cloudiness. By contrast, the RTNEPH has greater zones of virtually cloud free skies. There is better agreement between the three maps in the > 80 percent areas: over the western Mediterranean, northern Scandinavia, and northwest of the Caspian sea. All three of these general areas are also characterized by high (> 80 percent) probability of MDA-diagnosis cloud amount accuracy. However, there are areas of > 50 percent probability that do not correspond to agreement between the three cloud depictions: over the Black Sea, and over much of the Baltic Sea. Generally, however, the areas of > 80 percent probability correspond to areas of > 80 percent MDA-diagnosed cloud amount. This is most likely due to the fact that the 85-100 percent cloud amount category is the most prevalent category for total cloud.

7.3.2 0000 UTC 25 JANUARY 1991 EASTERN CENTRAL CHINA CASE

Figure 31 is a display of high cloud amount over eastern and central China as depicted by the transformed RTNEPH and 12-hour predictions using the MLR and MDA methods. Also shown are the probabilities of the MDA-predicted cloud amounts. Here we see high cloud limited to eastern-central China in all three cloud amount depictions. However, the areal extent of cloudiness is somewhat more limited in MLR than in RTNEPH, and even more limited in MDA. Both schemes underestimate the westward extent of the > 20 percent cloud amounts, and create a more concentrated area of > 50 percent cloud amounts than is apparent in the RTNEPH. As in the European case, the high cloud amount category probabilities from MDA are lowest in the areas where cloud amounts are diagnosed. This again reflects the dominance of 0 percent cloud amounts in the high cloud frequency of occurrence distribution, making 0 percent cloud a much more likely event. Across the southern section of the map and in the northeast sector, the > 80 percent

106
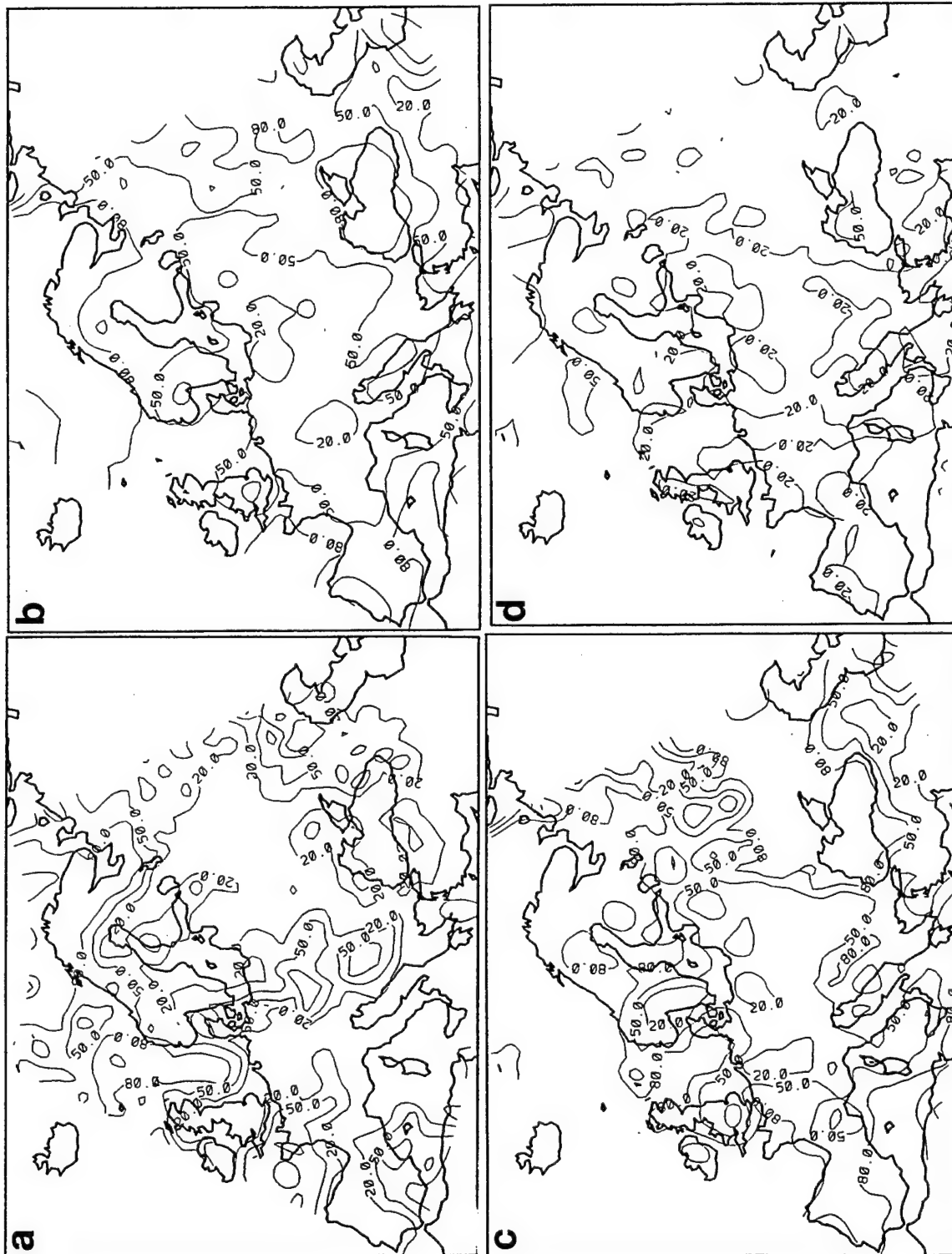
Figure 29. Same as in Figure 27 for Low Deck Cloud Amounts.
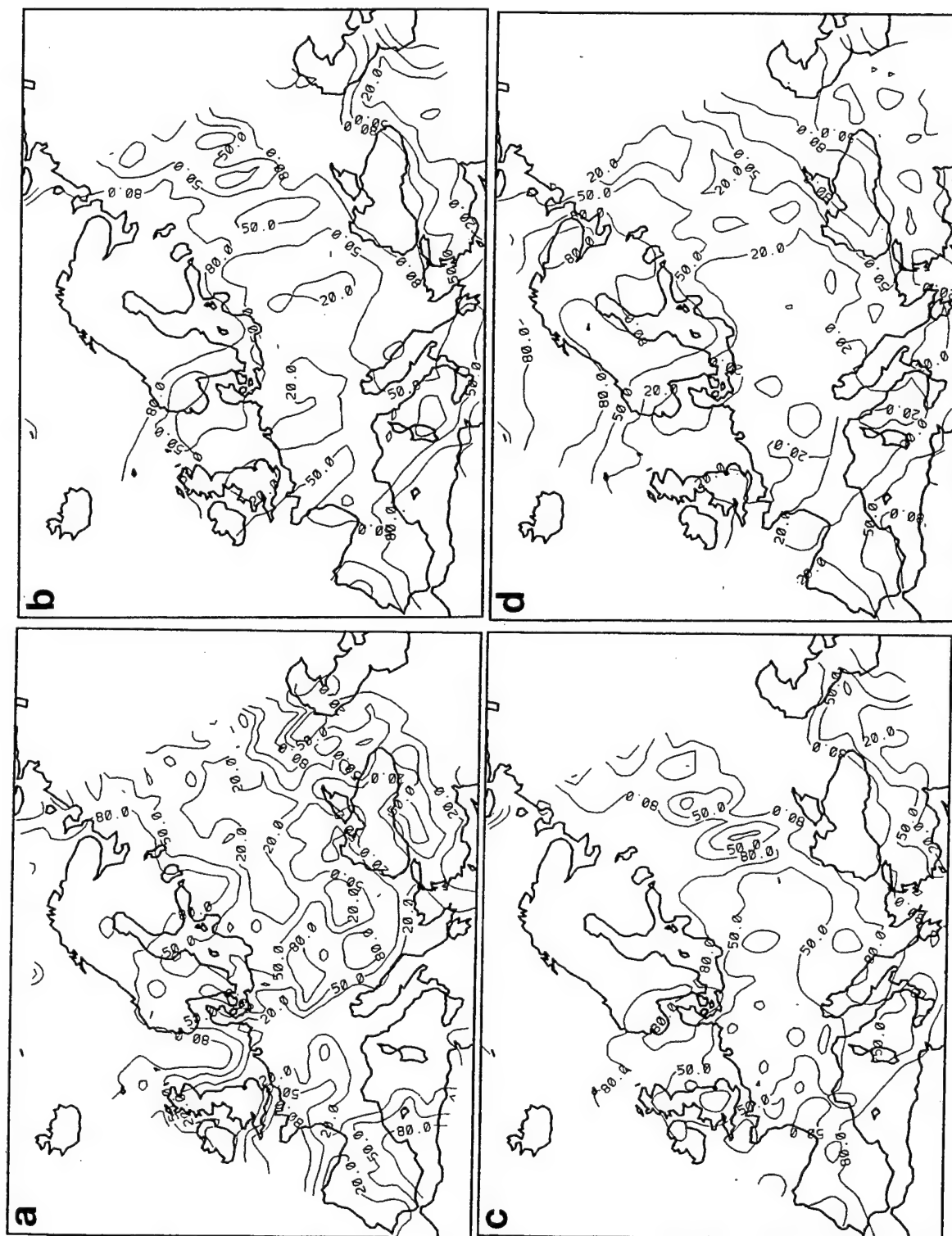
Figure 30. Same as in Figure 27 for Total Cloud Amounts.

108

probabilities of < 20 percent MDA cloud amount are consistent with the virtually cloud-free RTNEPH in these areas.

In the middle cloud deck depiction (Figure 32), there is again a similarity between the location of MLR-and MDA-diagnosed clouds but a much greater difference in magnitudes. Though both methods produce both a western and eastern bank of clouds, the MDA produces greater amounts in the western cloud mass. The MDA also produces a more westward extent to the eastern cloud mass. The RTNEPH verification depicts the cloudiness in the extreme eastern portion of the figure but depicts the largest area of cloudiness in the central portion of the figure between the MLR- and MDA-produced eastern and western cloud banks. As may be expected, the MDA probabilities are below 50 percent through the entire area occupied by clouds in any of the three depictions.

The transformed RTNEPH low cloud map (Figure 33) shows a large area of large cloud amounts in southeastern China. The gradients of cloud amount along the edges of the cloud mass are also large. These features are better reflected in the MDA diagnosis than in MLR's since the latter produces almost no areas of > 80 percent. However, both erroneously spread the clouds from the southeast continuously through to the northwest and over-represent the amounts in the latter portion of the map. Once again, the MDA probability map assigns lower probabilities (< 50 percent) to the areas where the MDA diagnosed the larger cloud amounts.

In Figure 34, we again see the value of a separate diagnosis of total cloud amount rather than stacking diagnosed deck cloud amounts. In both MLR and MDA, we see that total cloud maps do not have some of the erroneously large cloud amounts that we saw in the low cloud depictions in Figure 33--most notably in the extreme western portion of the maps. Both MLR and MDA total cloud depictions are clearly superior in accuracy to their low cloud counterparts. We have said previously that total cloud amounts are very probably more accurate than deck cloud amounts in the RTNEPH. This higher degree of accuracy may then lead to higher accuracy in the resulting total cloud diagnoses. Whereas both MLR and
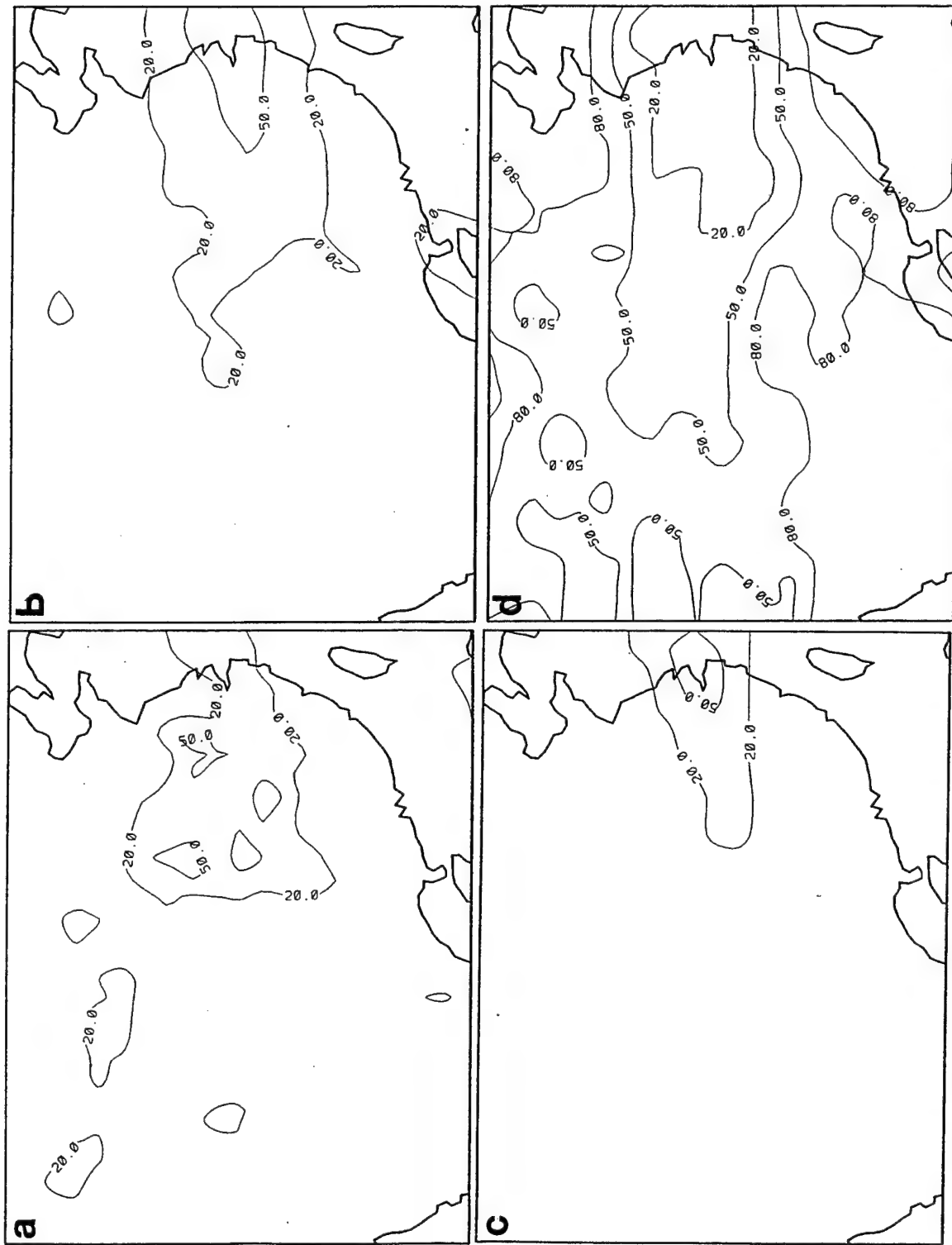
Figure 31. High Deck Cloud Amounts (%) Over Eastern Asia on 0000 UTC 25 January 1991 from (a) Transformed RTNEPH, (b) 12-Hour MLR Forecast, (c) 12-Hour MDA Forecast, and (d) 12-Hour MDA Forecast Probabilities.
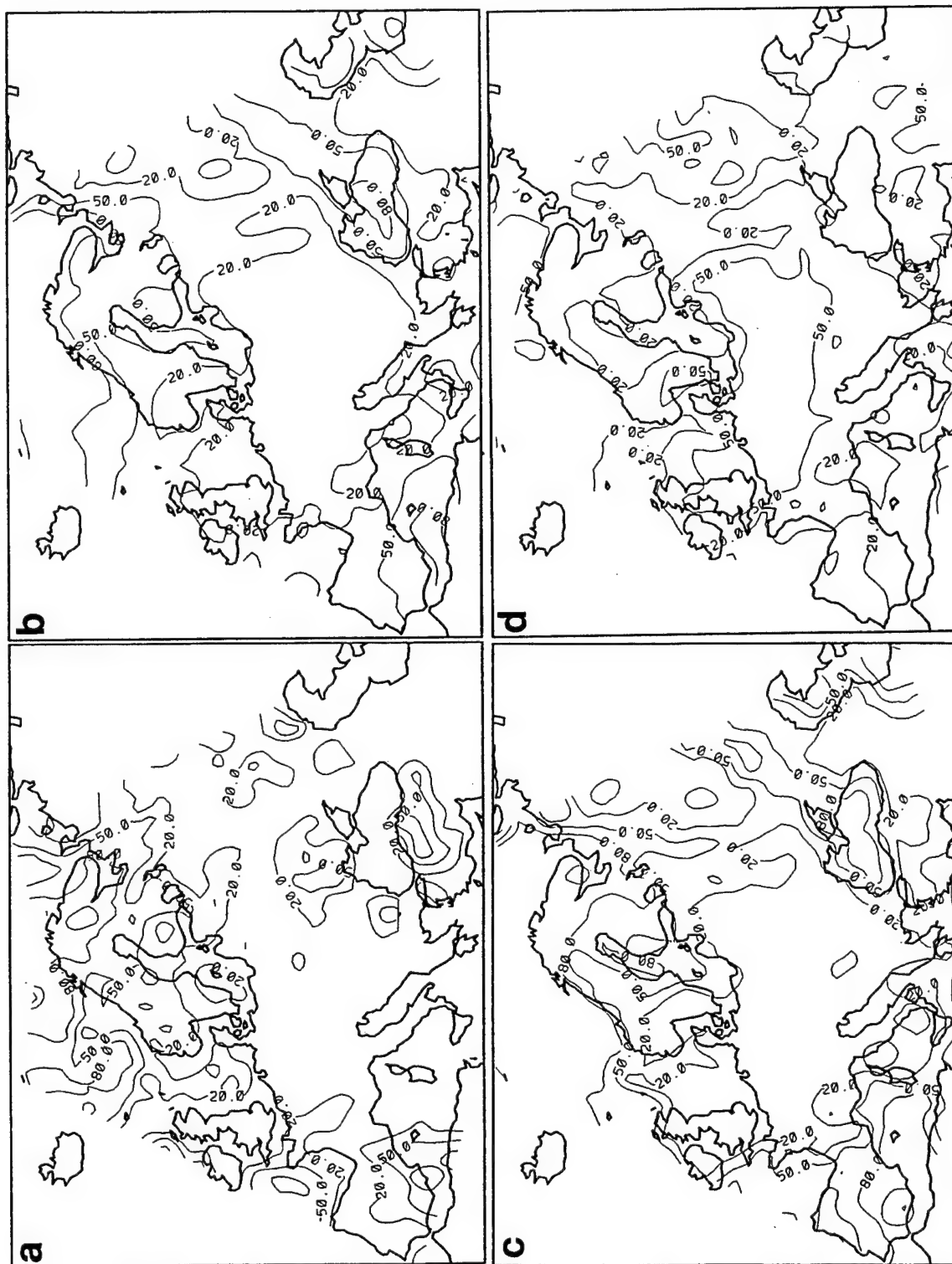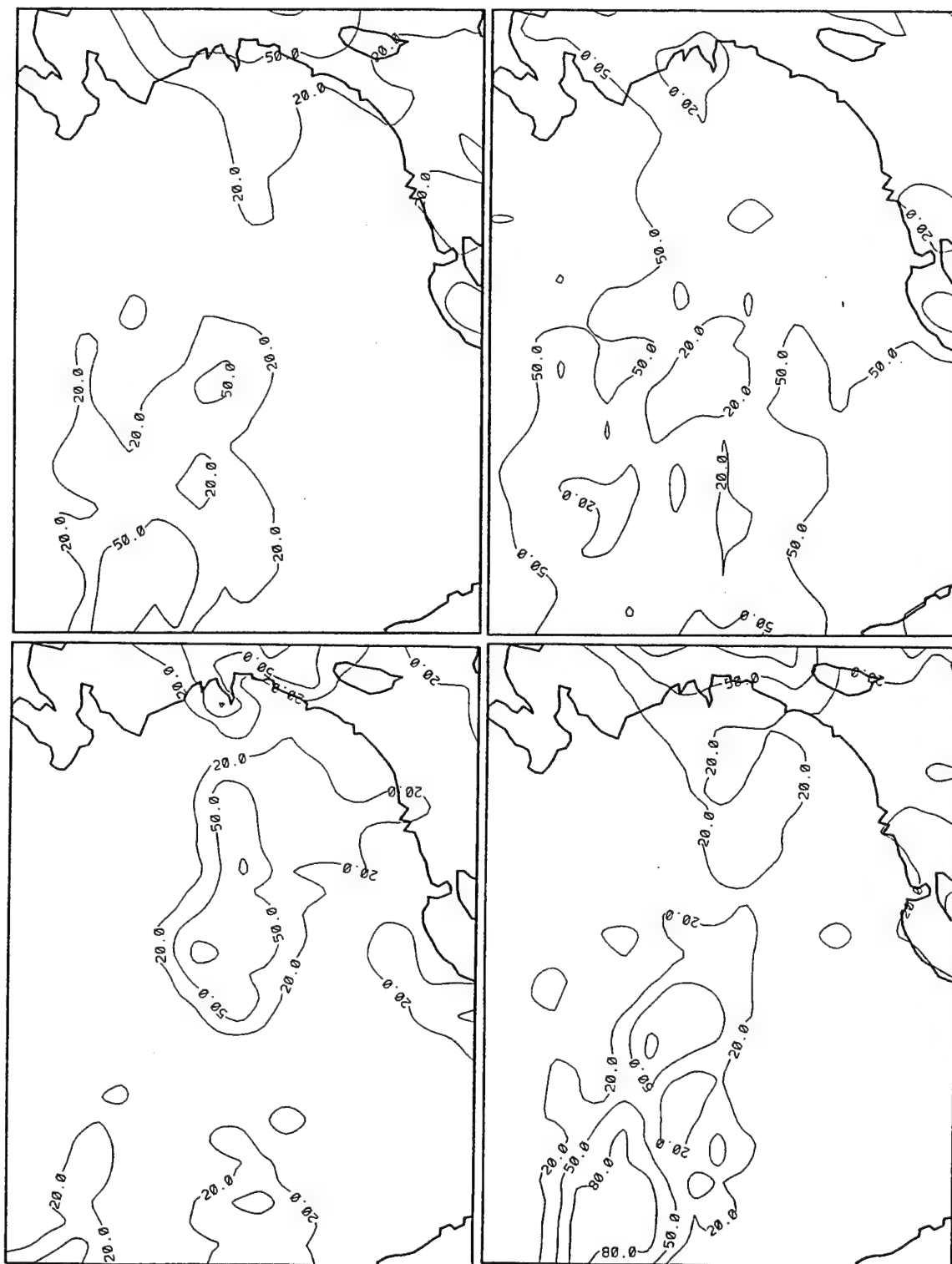
Figure 32. Same as in Figure 31 for Middle Deck Cloud Amounts.

111
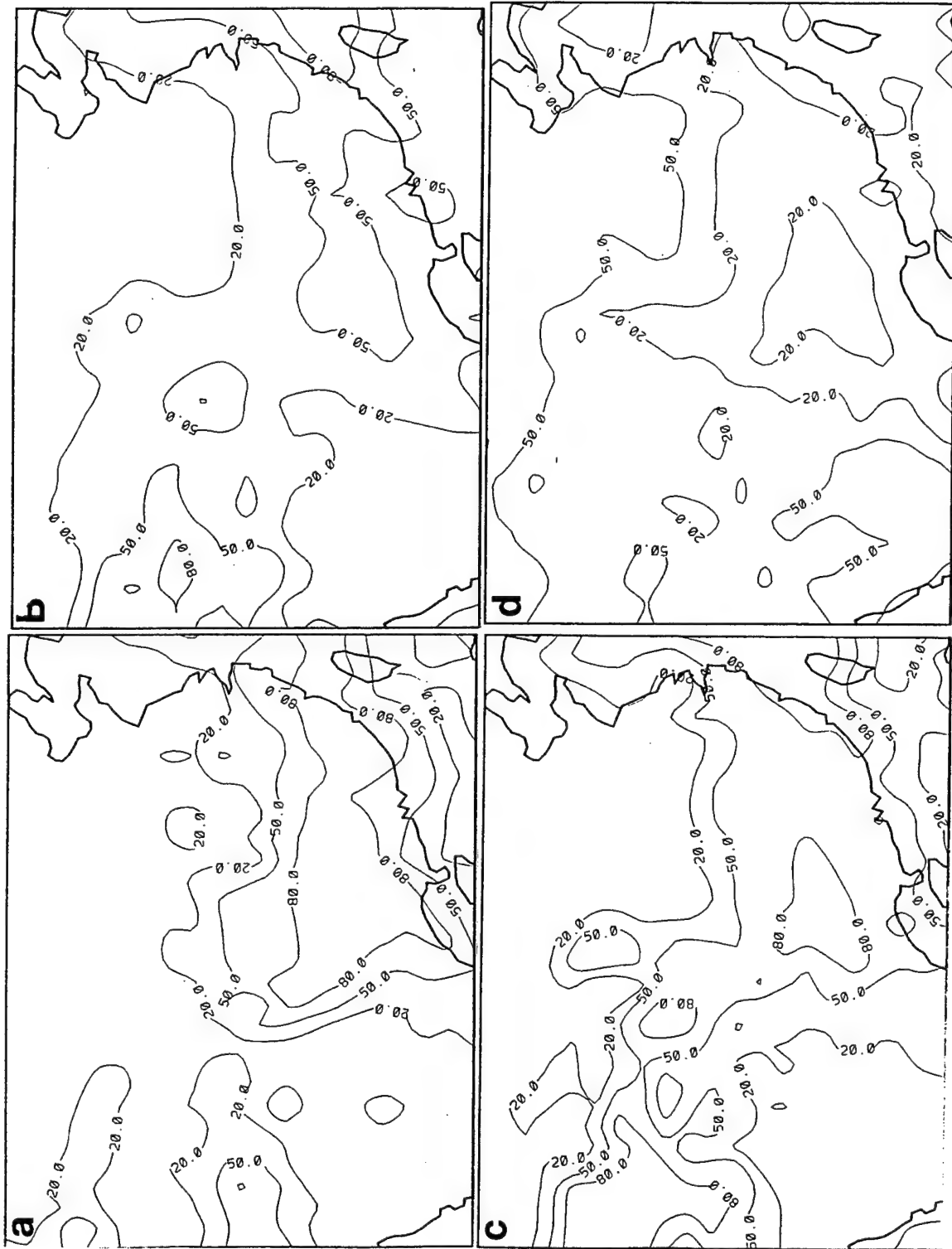
Figure 33. Same as in Figure 31 for Low Deck Cloud Amounts.

112
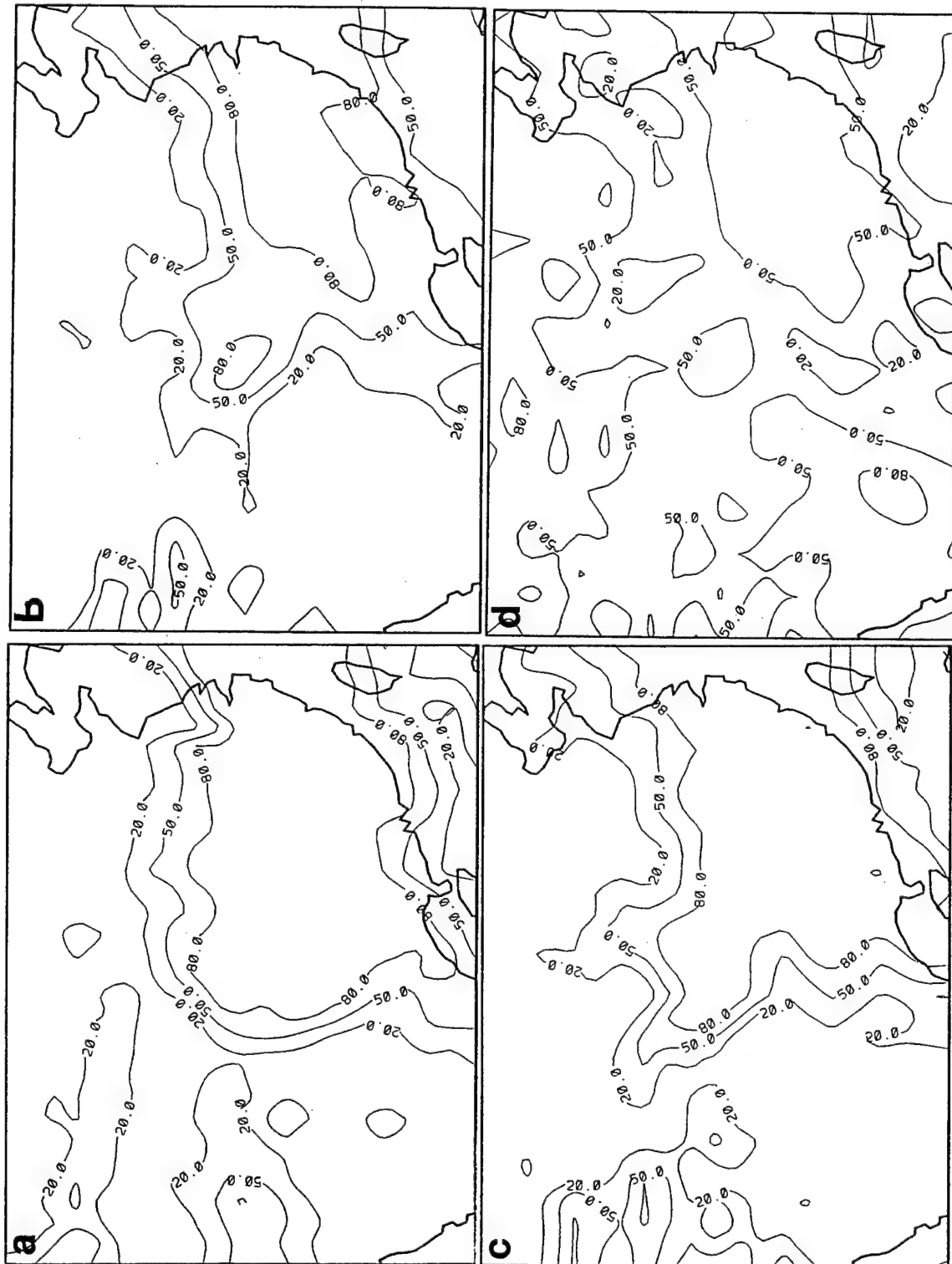
Figure 34. Same as in Figure 31 for Total Cloud Amounts.

MDA do display the > 80 percent cloud amounts in the southeast, the MDA depiction of the entire southwestern cloud mass is in much better agreement with the RTNEPH than is the MLR representation. The larger area of > 80 percent cloud amount in MDA results in the more realistic larger gradients along the edges of the cloud mass. Both methods properly separate the larger eastern cloud mass from the smaller western cloudy area, even though both methods overestimate cloud amounts in the latter area. Finally, we see that the largest probabilities (> 50 percent) of the MDA cloud amount category diagnosis lie in the areas of greatest agreement between MDA and RTNEPH cloud amounts. This is true in both the very cloudy southeastern area and in the virtually cloud-free southwestern, north central, and northeastern areas.

8. ADDITIONAL CLOUD DIAGNOSIS EXPERIMENTS

The foregoing experimentation was conducted to establish a baseline cloud diagnosis capability that could be used in operational forecast applications. We attempted to use predictands and predictors that would be available at AFGWC in an operational mode. The predictors were derived from a large-scale global NWP model that has model numerics, physics, and temporal and spatial resolutions that would be typically available from operational centers today. The predictands were derived from the currently operational AFGWC RTNEPH. We deliberately chose not to exploit data or techniques that would not be feasible for or available to an operational environment such as AFGWC. In this section, we remove this restriction and discuss the conduct of experiments designed to understand the implications of a change of spatial resolution, additional predictors, and an alternative predictand.

AFGWC is now preparing to receive large-scale global NWP model forecasts from the Navy's Fleet Numerical Oceanographic Center to provide global weather support needed for the development of any applicable AFGWC products. If the type of diagnostic cloud prediction procedures described in this report were executed

114

operationally at AFGWC, it would be this Navy product that would supply the weather forecast fields from which the predictors would be derived. These data are envisioned to be provided to AFGWC on 16 mandatory pressure levels on a 2.5° latitude-longitude grid. This stands in contrast to the predictors provided to the diagnosis procedures in this report: on 22 model sigma layers on a triangular 106 wave spectral coefficient form (an equivalent equal-area grid of 1.25° latitude-longitude). We thought that it would be of interest to examine the effect of the proposed coarser spatial resolution of the Navy's forecast fields on the performance of the cloud diagnosis assuming that the predictors derived from them would still be represented on the 1.25° latitude-longitude grid. The first of the experiments discussed below investigates this issue.

It is tempting to think that, if we could supply more discriminating information to the cloud diagnosis process, we may be able to improve the accuracy of the cloud amount diagnoses. Up to this point, we restricted the type of predictors supplied to the cloud diagnosis techniques to those that could be derived directly from the standard forecast field variables in a post-processing code executed at AFGWC. In our case, these were spectral coefficients of divergence, vorticity, temperature, and specific humidity on 22 model sigma layers, and the surface pressure. Also available were accumulated total precipitation, convective precipitation, and surface evaporation. All of these fields were available at 6-hour intervals, and from these fields we derived the 99 predictors listed in Table 1. At this point, we wished to investigate the utility of NWP model-diagnosed variables that are available only from the model itself. In particular, we wished to output at 6-h forecast intervals the values of potential supplemental predictors that are actually some of the diagnosed variables used in the NWP model's physical parameterization schemes. This would tell us the worth of attempting to make such variables available to the operational cloud diagnosis process at AFGWC. The second of the experiments discussed in this section attempts to answer this question.

Thirdly, we wished to try an alternative source of the predictand, cloud amount. As discussed in the introduction, the only known archived alternative cloud

analysis is the ISCCP[6,7] data set. Unfortunately, these data are available on an even coarser resolution than the AFGWC RTNEPH cloud analysis. However, a satellite-based experimental cloud cover analysis process known as SERCAA[23] has recently been developed as a part of a potential replacement for RTNEPH at AFGWC. At this time, SERCAA cloud amount analyses have been generated for only three limited geographic regions over approximately 10 days each. It was not clear if such a limited temporal-spatial sample of predictand would provide the necessary statistical robustness for the cloud diagnosis schemes described in this report. However, we decided to try the SERCAA cloud amount analyses as predictands in our methods, and the results are the product of our third experimental investigation described in this section.

8.1 Impact of Reduced Spatial Resolution of Predictors on Cloud Diagnosis

The goal of this experiment was to produce forecast cloud diagnoses from a reduced resolution version of the PL-94 forecast fields, and compare them with the diagnoses using the T106 resolution forecast fields described earlier in our report. By using exactly the same forecasts, predictors, and predictand, and changing only the spatial resolution of the forecast fields from which the predictors were derived, we could isolate the impact of just the coarser resolution on cloud diagnosis accuracy. For the purposes of this demonstration experiment, we concerned ourselves only with the predictors necessary for the 12-hour forecast diagnosis (the 6- and 12-hour forecast fields).

The first step was to reduce the horizontal resolution of the PL-94 T106 forecasts for the 17-day periods 8-24 January and July 1991. We did this by assigning a zero value to the wave number 48-106 spectral components of forecast divergence, vorticity, temperature, specific humidity and surface pressure. The resulting T47 spectral coefficients have an equivalent grid resolution that corresponds to a 2.5° latitude-longitude grid. The 22 layer T47 spectral coefficients (stored in the T106 spectral arrays) were then post-processed onto a 2.5° latitude-longitude grid, and

116

vertically interpolated onto the 16 mandatory pressure levels expected in the Navy forecast fields. The variables were geopotential height, temperature, wind components, and relative humidity fields. At this point, we had Navy forecast "look-alike" fields in structure, but from the PL-94 model. We had not yet reduced the resolution of the 1.125° Gaussian grid precipitation and evaporation fields from the model.

The next step was to introduce the gridded 2.5°, 16 level forecast fields into a T47L16 (16 layers chosen to approximate the distribution of mandatory pressure levels) version of our preprocessor. This module interpolates the pressure level data to the 16 model sigma layers and transforms the results in each model sigma to T47 spectral form. This was done to put the forecast data back in the form it has when it is produced by the forecast model: divergence, vorticity, temperature, specific humidity, and surface pressure.

Finally, the T47L16 spectral coefficient forecast fields, and the 1.125° latitude-longitude (original T106 forecast output) precipitation and evaporation data were introduced into our predictor generator module. The T47 spectral coefficients were evaluated on the original 1.125° equal-area grid, and used in the same way as the T106 forecasts to generate the predictor values on this grid. This included the vertical averaging of the predictor variables to high, middle, and low cloud decks as before.

We performed a Fourier transformation of the 1.125° longitude-spaced precipitation and evaporation values on each T106 Gaussian latitude to 47 waves to spatially truncate these data, then a backward transform onto the equal-area gridpoints. The output from the predictor generator module was then identical in form to the equal-area grid predictors obtained from the PL-94 T106L22 forecasts discussed earlier in this report.

We then performed 12-hour forecast cloud diagnoses using both MLR and 6-category MDA using these T47L16 predictors for the 7-day verification periods (days 18-24) in each of January and July of 1991. We used the 10-day period just prior to each verification date as the development period for each date that marks

the initial times of the 12-hour forecasts on which to perform the diagnosis of cloud amount. We then verified the resulting diagnoses against the transformed RTNEPH at the forecast-valid times. Because we used both MLR and MDA, we present just the cloud amount category verification in Table 25. We have included the T106L22 predictor results from Table 24 as a comparison to determine the impact of using the lower resolution predictors.

The comparison of the skill scores from the two different forecast field resolutions shows that skill was not compromised as a result of the truncation from T106L22 to T47L16. If anything, there is a slight improvement in skill resulting from the truncation, although the difference in skill may not be statistically significant. The slight improvement is evident in all decks, both months, and both diagnostic methods, with few isolated exceptions.

We looked at the total correlations of cloud amount with the linear combination of the predictors selected for three of the seven days in the verification period. We found that in all cases, the total correlation was slightly higher for the T47L16 diagnosis than for the T106L22 diagnosis. This finding is consistent with that of Walcek,[24] who found that the spatial correlation between cloud amount and various predictors improves with the size of the area over which the quantities are averaged to form the grid values. In our case, the spectral truncation effectively removes the shorter wavelength spatial variations in the predictor fields. This has the same effect as averaging finely gridded fields spatially to get a coarser grid representation. Because some of the shorter wavelength features are removed from the NWP forecasts, the correlations of the NWP predictors with cloud amount (averaged to the same grid resolution) improve. This may contribute toward slightly improved cloud amount diagnosis when cloud amount/predictor relationships are applied to independent predictor values.

8.2 Impact of Using Supplemental Predictors on Cloud Diagnosis

The predictors given in Table 2 are derived directly from the standard global spectral model outputs. These include model sigma level values of divergence,

Table 25. Comparison of Category Skill Scores Between T106L22 and T47L16 Forecast Predictors in 12-Hour Forecasts for the Verification Periods 18-24 January and July 1991, Northern Hemisphere

| | January | | | | July | | | |
| | MLR | | MDA | | MLR | | MDA | |
Deck	T47L16	T106L22	T47L16	T106L22	T47L22	T106L22	T47L16	T106L22
				Bias				
H	0.278	0.273	-0.016	-0.025	0.357	0.345	-0.093	-0.079
M	0.062	0.090	0.172	0.177	0.428	0.213	-0.050	-0.047
L	0.089	0.092	0.074	0.062	0.169	0.158	0.028	0.064
T	-0.146	-0.133	0.050	0.058	-0.027	-0.048	-0.070	-0.068
				RMSE				
H	1.346	1.354	1.392	1.402	1.523	1.521	1.547	1.559
M	1.694	1.703	1.874	1.893	1.792	1.796	1.976	1.982
L	1.691	1.702	1.889	1.907	1.760	1.774	1.972	1.973
T	1.706	1.719	1.860	1.866	1.664	1.681	1.743	1.767
				MAE				
H	0.851	0.863	0.735	0.743	0.991	0.997	0.885	0.897
M	1.258	1.268	1.333	1.345	1.352	1.358	1.441	1.446
L	1.257	1.264	1.331	1.342	1.348	1.360	1.454	1.459
T	1.211	1.217	1.241	1.242	1.101	1.122	1.132	1.150
				Percent in Correct Category				
H	47.4	46.9	62.0	61.5	44.1	43.4	55.5	55.0
M	29.0	28.6	32.9	33.0	26.4	26.2	29.6	29.5
L	29.4	29.5	34.6	34.5	25.1	24.6	28.7	28.3
T	34.9	34.8	40.7	40.7	42.6	41.5	44.4	43.9
				Frequency of Occurrence Fit				
H	0.530	0.537	0.041	0.041	0.412	0.431	0.058	0.052
M	0.406	0.419	0.088	0.081	0.428	0.426	0.069	0.068
L	0.379	0.378	0.047	0.061	0.526	0.520	0.063	0.068
T	0.353	0.359	0.037	0.041	0.219	0.255	0.094	0.086

vorticity, temperature, and specific humidity, and pressure at the model terrain surface. In addition, grid values of cumulative total precipitation, convective precipitation, and evaporation amounts are included in the standard outputs.

We wished to determine how much additional benefit could be provided to the cloud diagnosis process by modifying the global spectral model to output supplemental information. We felt that there were a number of quantities diagnosed within the model that might have potential to discriminate cloud amount distributions more fully. When used in conjunction with some of the variables we had already identified, we wished to see if these additional variables might boost the total correlation between predictand and predictors.

Our first step was to eliminate any of the current predictors that were not making a contribution to the correlation with cloud amount. Because we had already supplied 99 predictors to the regression algorithm of which only the top 20 were selected, we saw no need to continue to carry non-contributors when adding supplementary variables. To identify the non-contributors, we examined the 20 predictors selected in each deck by the forward stepwise regression in the cloud diagnosis executions of MLR for 8-17, 11-20, and 14-23 January and July in both 12- and 48-hour forecasts. Any of the 99 predictors that were not selected more than once over all of these cases were eliminated. If any of the deck values of a predictor was selected more than once, all deck values of that predictor were retained. On this basis, the following predictors were eliminated from the set of 99 predictors (see Table 2): 2, 6, 10, 11, 22, 29, 30, 31, 52, 53, 72, 83, 85, 86, 93, 94. In addition, we replaced surface pressure with mean sea level pressure at both forecast times t-6 and t.

We next went through the PL GSM code to determine which diagnostic variables might be likely candidates for supplemental cloud amount predictors. The variables chosen as supplemental predictors are listed in Table 26 as numbers 61-82. We concentrated our attention solely on the physics parameterizations in the model: radiation (61 and 62), planetary boundary layer (63-77), gravity wave drag (78), and cumulus convection (79-82). We modified the PL-94 version of the PL

GSM to output these quantities on the Gaussian grid (in this case, 320 gridpoints on each Gaussian latitude) at each 6-hour forecast time interval out to 48 hours. We then reran the PL-94 model to output both these fields and the standard forecast output as before.

Our modified version of the predictor file generator then used these data as input and generated the 104 predictor variables shown in Table 26 on the equal-area grid for each cloud deck. For each supplemental predictor on each Gaussian latitude, we performed a Fourier transform on the 1.125° longitude-spaced values output by PL-94, then evaluated the Fourier coefficients on the equal-area gridpoints. The standard forecast outputs were processed in a manner identical to that used for the list of original predictors. The geographic predictors that were retained in the new predictor list had values identical to those used in the original list. The net result was a list of predictors that did not include non-contributing predictors but did include new supplemental predictors.

We used the predictor values as input to both the MLR and MDA cloud diagnosis methods for both 12- and 48-hour forecasts. Because we had added the supplemental predictors, we were interested in how many of them were selected in the top 20 predictors in the forward stepwise regression process of the MLR. Also of importance are the predictors in the original list that are now not selected in the top 20, having been displaced by the new predictors. The predictors that were evaluated as strong or useful by the forward stepwise selection method applied to the new list of predictors using 12-hour forecast values are given in Table 27. These may be compared with predictors selected from the original list of predictors shown in Table 3.

We can compare the two tables to determine to what extent the supplemental predictors displaced the original predictors. The original predictor list variables that were selected less in the new list in January were: temperature, meridional wind, sine(longitude), surface terrain standard deviation, $3 \times 3 \times 3$ maximum wind speed, 3×3 maximum surface wind speed, evaporation rate, and surface layer wind speed. The supplemental predictors that were selected most in January were

Table 26. List and Description of the "105-Predictors" Used in the Selection of Multi-Linear-Regression Cloud Predictors.

No.	Name	Description
1	VORD6	Vorticity, predictand deck average, forecast t-6
2	TMPD6	Temperature, predictand deck average, forecast t-6
3	PRWD6	Precipitable water, predictand deck average, forecast t-6
4	RHUD6	Relative humidity, predictand deck average, forecast t-6
5	STBD6	d(theta)/d(z), predictand deck average, forecast t-6
6	SPDD6	Wind speed, predictand deck average, forecast t-6
7	SHRD6	Wind shear, predictand deck average, forecast t-6
8	QADD6	3-D humidity div., predictand deck average, forecast t-6
9	CPSD6	Condens. pres. deficit, predictand deck average, forecast t-6
10	MSTD6	d(theta-e)/d(z), predictand deck average, forecast t-6
11	UCMD6	West wind component, predictand deck average, forecast t-6
12	VCMD6	South wind component, predictand deck average, forecast t-6
13	RHXC6	Maximum RH within predictand deck, forecast t-6
14	RHAC6	RH at layer above maximum RH (see #13), forecast t-6
15	TMPC6	Temperature at maximum RH (see #13), forecast t-6
16	STBC6	d(theta)/d(z) at maximun RH (see #13), forecast t-6
17	MSLP6	Sea level pressure, forecast t-6
18	RFCV6	6-hr convective surface precipitation, forecast t-6
19	EVAP6	6-hr surface evaporation, forecast t-6
20	SPDB6	Surface-layer wind speed, forecast t-6
21	VORH2	Vorticity, high deck average, forecast t-0
22	VORM2	Vorticity, middle deck average, forecast t-0
23	VORL2	Vorticity, low deck average, forecast t-0
24	RHUH2	Relative humidity (RH), high deck average, forecast t-0
25	RHUM2	Relative humidity, middle deck average, forecast t-0
26	RHUL2	Relative humidity, low deck average, forecast t-0
27	OMGH2	Vertical velocity, high deck average, forecast t-0
28	OMGM2	Vertical velocity, middle deck average, forecast t-0
29	OMGL2	Vertical velocity, low deck average, forecast t-0
30	STBH2	d(theta)/d(z), high deck average, forecast t-0
31	STBM2	d(theta)/d(z), middle deck average, forecast t-0
32	STBL2	d(theta)/d(z), low deck average, forecast t-0
33	SPDH2	Wind speed, high deck average, forecast t-0
34	SPDM2	Wind speed, middle deck average, forecast t-0
35	SPDL2	Wind speed, low deck average, forecast t-0
36	SHRH2	Wind shear, high deck average, forecast t-0
37	SHRM2	Wind shear, middle deck average, forecast t-0
38	SHRL2	Wind shear, low deck average, forecast t-0
39	RHCH2	Maximum RH within high deck, forecast t-0
40	RHCM2	Maximum RH within middle deck, forecast t-0
41	RHCL2	Maximum RH within low deck, forecast t-0
42	TMPD2	Temperature, predictand deck average, forecast t-0
43	PRWD2	Precipitable water, predictand deck average, forecast t-0
44	QADD2	3-D humidity div., predictand deck average, forecast t-0
45	CPSD2	Condens. pres. deficit, predictand deck average, forecast t-0
46	MSTD2	d(theta-e)/d(z), predictand deck average, forecast t-0
47	UCMD2	West wind component, predictand deck average, forecast t-0
48	VCMD2	South wind component, predictand deck average, forecast t-0
49	RHAC2	RH for level above RH-max, predictand deck, forecast t-0
50	TMPC2	Temperature at maximum RH (see #39-41), forecast t-0
51	STBC2	d(theta)/d(z) at maximum RH (see #39-41), forecast t-0
52	MSLP2	Sea level pressure, forecast t-0

Table 26. (cont.) List and Description of the "105-Predictors" Used in the Selection of Multi-Linear-Regression Cloud Predictors.

No.	Name	Description
53	RFST2	6-hr stratiform surface precipitation, forecast t-0
54	RFCV2	6-hr convective surface precipitation, forecast t-0
55	EVAP2	6-hr surface evaporation, forecast t-0
56	SPDB2	Surface-layer wind speed, forecast t-0
57	RH2C2	Maximum-RH-squared within predictand deck, forecast t-0
58	RH4C2	Maximum-RH-fourth within predictand deck, forecast t-0
59	RHIC2	RH wrt ice at RH maximum, predictand deck, forecast t-0
60	LCDC2	Lifted-cond.-dist. at RH maximum, pred. deck, forecast t-0
61	RADT2	GSM radiation temperature tendency, pred. deck ave, forecast t-0
62	SFRF2	GSM net downward surface radiative flux, forecast t-0
63	SRLM2	GSM surface roughness length - momentum, forecast t-0
64	SRLH2	GSM surface roughness length - heat, forecast t-0
65	SRLQ2	GSM surface roughness length - moisture, forecast t-0
66	BRIN2	GSM bulk Richardson number, forecast t-0
67	SECM2	GSM surface exchange coefficient - momentum, forecast t-0
68	SECH2	GSM surface exchange coefficient - heat, forecast t-0
69	SSPH2	GSM surface specific humidity, forecast t-0
70	STMP2	GSM surface temperature, forecast t-0
71	RMOL2	GSM reciprocal of Monin-Obukhov length, forecast t-0
72	FRIN2	GSM flux Richardson number, forecast t-0
73	PBLH2	GSM planetary boundary layer (PBL) height, forecast t-0
74	TBDC2	GSM turbulence diffusion coeff., pred. deck ave., forecast t-0
75	PWST2	GSM PBL wind speed tendency, pred. deck ave., forecast t-0
76	PTMT2	GSM PBL temperature tendency, pred. deck ave., forecast t-0
77	PSHT2	GSM PBL specific humid. tendency, pred. deck ave., forecast t-0
78	GWDT2	GSM grav. wave drag wind spd tend, pred. deck ave., forecast t-0
79	CCLW2	GSM convective cloud liquid water, pred. deck ave., forecast t-0
80	CWST2	GSM convective wind spd tendency, pred. deck ave., forecast t-0
81	CTMT2	GSM convective temperature tend., pred. deck ave., forecast t-0
82	CSHT2	GSM convective spec. humid. tend., pred. deck ave., forecast t-0
83	GSLAT	Latitude (Gaussian grid, GS)
84	SGSLA	Sine of Latitude
85	CGSLA	Cosine of Latitude
86	SGSLO	Sine of Longitude
87	CGSLO	Cosine of Longitude
88	ZENA2	Solar zenith angle, forecast t-0
89	CZEN2	Cosine of solar zenith angle, forecast t-0
90	HRSS2	Hours of sunshine before forecast t-0
91	HRDK2	Hours of darkness before forecast t-0
92	SFCHT	Surface terrain height (9-pt ave., 1/8 mesh data)
93	PCH2O	Percent of surface that is water (from 1/64 mesh data)
94	LRN92	3x3x3 (ijk) minimum of Ln(Ri.-Number), forecast t-0
95	STN92	3x3x3 minimum of d(theta)/d(z), deck average, forecast t-0
96	SHX92	3x3x3 maximum of vertical shear, deck average, forecast t-0
97	SPX92	3x3x3 maximum of wind speed, deck average, forecast t-0
98	RCX92	3x3 maximum of 6-hr convective rainfall, forecast t-0
99	SBX92	3x3 maximum of surface layer wind speed, forecast t-0
100	ABTV2	Min. of stnd dev terrain ht. or wind/stability height, frcst t-0
101	RH2D2	RH-squared, predictand deck average, forecast t-0
102	RH4D2	RH-fourth, predictand deck average, forecast t-0
103	RH2D6	RH-squared, predictand deck average, forecast t-6
104	RH4D6	RH-fourth, predictand deck average, forecast t-6
105	CLDOB	Predictand, observed RTNEPH deck cloud cover, forecast t-0

planetary boundary layer height and surface specific humidity. No other January supplementary variables were more than useful predictors in one or two cloud decks. The original predictor list variables that were selected less in the new list in July were: temperature, relative humidity, precipitable water, sine(longitude), cosine of zenith angle, hours of both sunlight and dark before forecast time, surface height, zenith angle, wind shear, $3 \times 3 \times 3$ maximum wind speed, 3×3 maximum surface wind speed, and surface layer wind speed. The supplemental predictor that was selected most in July was radiation temperature tendency.

In addition to the supplemental predictors, we replaced surface pressure in the original list with sea level pressure in the new list. This replacement effectively removes the terrain influence from the surface atmospheric pressure distribution. Thus, we might expect that the contribution of terrain height (called surface height in Tables 3 and 27) may be influenced in some way.

In comparing Tables 3 and 27, we in fact find very little change in the degree of contribution in either the pressure or the height predictors in January. Perhaps the only significant change is the reduction from strong to useful in changing from surface to sea level pressure in the low deck. However, the increase in contribution in changing from surface to sea level pressure couldn't be greater in July. We find that sea level pressure is now a strong predictor in all four cloud decks, whereas surface pressure was a non-contributor. Surface height decreased in its contribution to the total correlation in switching from the old to the new predictor list. However, we can't necessarily attribute this change to the change from surface to sea level pressure, because many other variables were changed between the old and new predictor lists as well.

A comparison of the total correlation of the predictors selected from the old and new lists with cloud amount showed that the correlations with the new predictors were insignificantly higher (not more than 3 percent) in all four cloud decks in both months. The results of the 12-h forecast cloud amount diagnosis verifications for both January and July 7-day verification periods are shown in Table 28. In the cloud amount diagnosis skill score comparisons, we find no consistent nor

124

Table 27. Strong (x) and Useful (+) Predictors in 12-Hour PL-94 Forecasts Using 105-Predictor List

JANUARY

Dynamic

Dynamic	L	H	T
temperature	+		x
zonal wind		+	
meridional wind			x
sea lvl pressure	+		+
temp @ max RH lvl			x
sfc temperature	+	+	
sfc spec humid	x		

Humidity

Humidity	L	M	H	T
relative humidity	x#	x⁻	x	x⁻
(rel humid)²	x		x	x
(rel humid)⁴	x		+	x
precip water	+		x	+
ice rel humid max	+			x
cond pres spread	+		+	
lifting cond level	x			
RH above max RH		+		
(rel humid max)²	+			
rel humid max				+

Geographic

Geographic	L	M	H	T
sin (longitude)	x#	x⁻	x	x⁻
cos (longitude)	x		x	x
hrs. sun before t	x		+	x
hrs. dark before t	+		x	+
% surface water	+			+
cos (latitude)	+		+	x
surface height	x			

Turbulence

Turbulence	L	M	H	T
dry static stab				+
moist static stab	+	+		x
wind speed	+*	+		
wind shear				+
3X3X3 min stab	+	x	x	x
3X3 sfc max speed	+			x
evaporation rate		+		
conv prep rate		+	+	
rad temp tend		+	+	
conv cloud water		+	+	+
sfc rad flux	+	+		
PBL height	+	+	+	
Z* min/wind-stab	+			x

JULY

Dynamic

Dynamic	L	M	H	T
temperature	x	x*@	+@#	x
meridional wind	+	+	+	x
stratiform prcp		x	x	+
sea lvl pressure	x	x@#	x@#	x#

Humidity

Humidity	L	M	H	T
relative	+	x*@	+@#	x
(rel humid)²	+	+	+	x
(rel humid)⁴		x	+	+
precip water	x			x
cond pres	x	+	+	x
rel humid max	x#	x@#	+	x#
RH above max		+	+	+

Geographic

Geographic	L	M	H	T
sin (latitude)	+			x
latitude	+			x
sin (longitude)		x		+
cos (longitude)	x			+
hrs dark before t	+			x
% surface water	+	x	x	x#
surface height	+		+	+
zenith angle	+			
cos (latitude)				+

Turbulence

Turbulence	L	M	H	T
dry static stab	+			+
wind speed	+		x	x
wind shear		x		
evaporation rate	x	x	x	x
sfc rough len-q		+		
rad temp tend	x		x	+
PBL temp tend	x			+
3X3X3 max shear	x			+

Predictors are from the deck of diagnosis except as follows:
* low deck; @ middle deck; # high deck; ~all 3 decks

Note: No distinction is made here between t-6 and t-0 values

significant difference in cloud amount prediction skill between the use of the old and the new predictors. It appears that no advantage in cloud amount prediction is gained through the use of the supplemental predictors that we chose to use in this study.

8.3 Use of SERCAA Total Cloud Amount as the Predictand

The SERCAA[23] layered and total cloud cover products are composed entirely from a combination of separate cloud analyses from several satellite system imagers. These include the optical line scan (OLS) imager on polar-orbiting Defense Meteorological Satellite Program satellites, the advanced very high resolution radiometer (AVHRR) aboard the polar-orbiting National Oceanic and Atmospheric Administration (NOAA) satellites, the visible infrared spin-scan radiometer (VISSR) on the geostationary Meteorological Satellite-Europe (METEOSAT) and Geostationary Meteorological Satellite-Japan (GMS) satellites, and the VISSR atmospheric sounder (VAS) on board the geostationary Geosynchronous Operational Environmental Satellite-U.S. (GOES) satellite. Once the layered and total cloud cover products are derived from each satellite type listed above, the SERCAA integration algorithm combines the separate analyses into a combined analysis. It is the combined analysis product that we chose to use in this study. As previously stated, the SERCAA cloud cover product is based on satellite data only. It uses no conventional observations as does RTNEPH. As such, thecloud cover percentages represented in the SERCAA analyses are those that can be seen by satellite. This means that the layered cloud amounts reported are the portions of the layer clouds not obscured by higher clouds. As a result, the layer cloud cover reported would not necessarily include all of the cloud in the layers if there was part of the layer cloud obscured by cloud above. For this reason, we used only total cloud cover in this study.

Table 28. Comparison of Category Skill Scores Between Old and New Predictor Lists in 12-Hour Forecasts for the Verification Periods 18-24 January and July 1991, Northern Hemisphere.

| | January | | | | | July | | | |
| | MLR | | MDA | | | MLR | | MDA | |
Deck	Old	New	Old	New		Old	New	Old	New
				Bias					
H	0.273	0.282	-0.025	-0.027		0.345	0.357	-0.079	-0.099
M	0.090	0.068	0.177	0.197		0.213	0.199	-0.047	0.031
L	0.092	0.094	0.062	0.060		0.158	0.141	0.064	0.099
T	-0.133	-0.145	0.058	0.048		-0.048	-0.048	-0.068	-0.056
				RMSE					
H	1.354	1.358	1.402	1.397		1.521	1.523	1.559	1.551
M	1.703	1.708	1.893	1.909		1.796	1.806	1.982	2.010
L	1.702	1.703	1.907	1.914		1.774	1.782	1.973	1.983
T	1.719	1.718	1.866	1.873		1.681	1.698	1.767	1.762
				MAE					
H	0.863	0.866	0.743	0.740		0.997	0.999	0.897	0.889
M	1.268	1.275	1.345	1.363		1.358	1.370	1.446	1.476
L	1.264	1.267	1.342	1.347		1.360	1.361	1.459	1.465
T	1.217	1.216	1.242	1.250		1.222	1.127	1.150	1.143
				Percent in Correct Category					
H	46.9	46.8	61.5	61.6		43.4	43.4	55.0	55.4
M	28.6	28.4	33.0	32.2		26.2	25.9	29.5	29.0
L	29.5	29.3	34.5	34.5		24.6	25.2	28.3	28.5
T	34.8	35.0	40.7	40.6		41.5	41.9	43.9	44.2
				Frequency of Occurrence Fit					
H	0.537	0.542	0.041	0.044		0.431	0.434	0.052	0.062
M	0.419	0.448	0.081	0.130		0.426	0.425	0.068	0.046
L	0.378	0.391	0.061	0.069		0.520	0.491	0.068	0.069
T	0.359	0.350	0.041	0.040		0.255	0.223	0.086	0.085

127

The SERCAA cloud cover analyses are available on a horizontal grid of resolution 12.5 nautical miles (about 23 km) true at 60 degrees latitude. In fact, it uses exactly the same grid as the RTNEPH but with twice as many gridpoints in both I and J directions. This required that we now use a 5×5 average of SERCAA gridpoint cloud cover values centered on the SERCAA gridpoint lying closest to each equal-area gridpoint. We used all of the available SERCAA cloud cover values without regard to their timeliness. In fact, with the use of a constellation of satellites in building the SERCAA analysis, timeliness was considered in the blending of the separate satellite analyses into the combined SERCAA analyses.

SERCAA cloud cover analyses have been constructed for limited time periods over several limited-area regions of the Northern Hemisphere. In Figure 35, we show the two regions of interest used in this study--the eastern Mediterranean (EMD) region and the central America (CNS) region. Analyses were constructed for four times daily (00, 06, 12, 18 UTC) for an 11-day period (12-22 March 1994) over EMD and a 9-day period (24 March-1 April 1994) over CNS. We transformed the total cloud cover values on the SERCAA grid to our equal-area grid in a manner directly analogous to the transformation of the RTNEPH cloud amount analyses, except in this case we used a 5×5 array of SERCAA gridpoints for the equal-area gridpoint cloud cover value. As before, we rounded the resulting transformed grid cloud cover value to the nearest 5 percent cloud cover. In this way, we constructed equal-area grid values of total cloud cover for use as the predictand over the EMD and CNS regions for their respective 11- and 9-day periods. The frequency distribution of the total cloud cover for both regions is shown in Figure 36.

To develop the predictors, we obtained twice-daily global meteorological analysis data sets for the period 11-31 March 1994. These data sets were generated by the National Centers for Environmental Prediction (NCEP) Spectral Statistical Interpolation (SSI) analysis model. We acquired the data sets from the National Center for Atmospheric Research (NCAR). The data came in the form of triangular 126 wave fields of temperature, vorticity, divergence, and specific humidity on 28 model sigma layers, and surface pressure and terrain height. We evaluated the
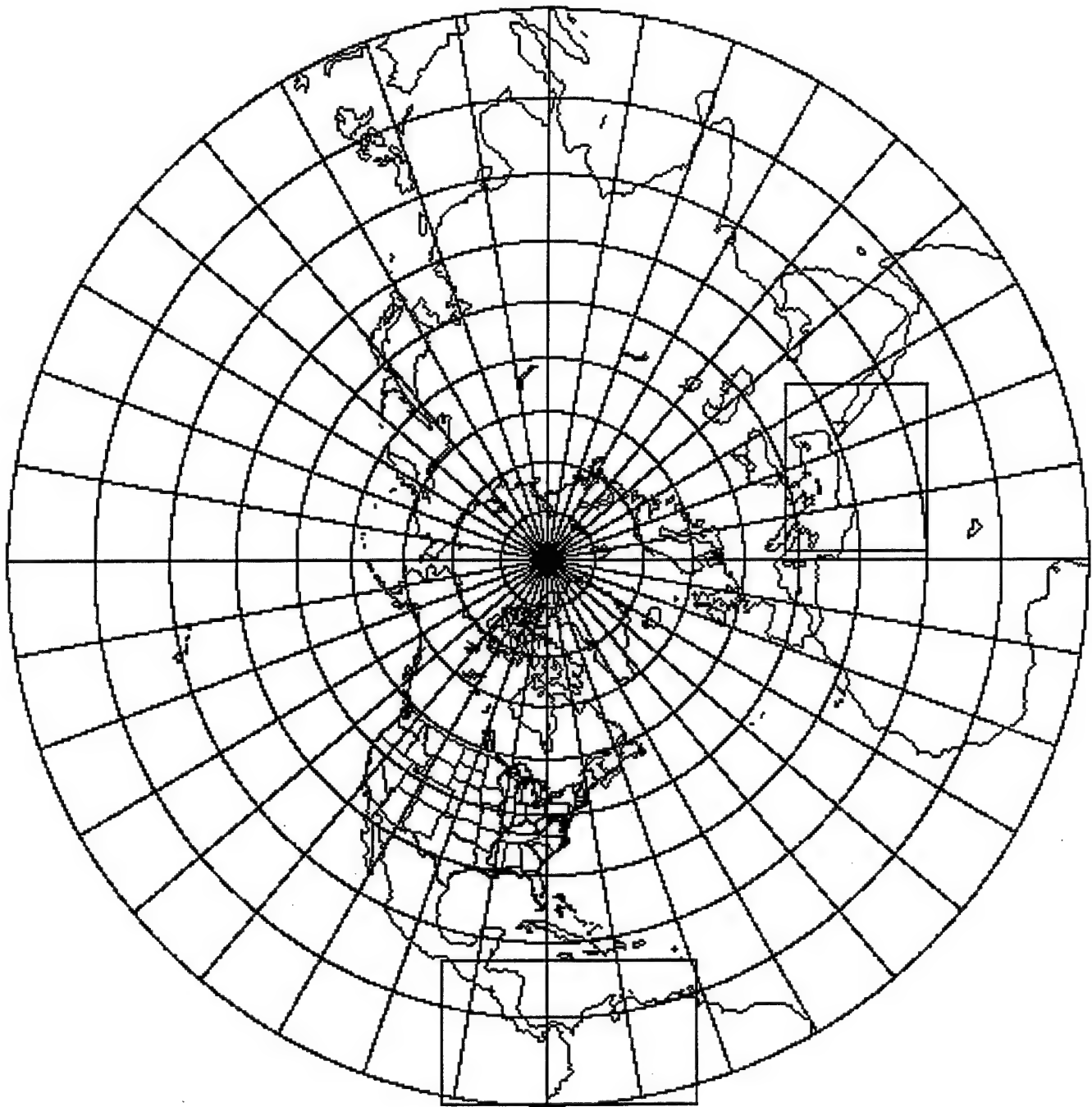
128

Figure 35. SERCAA Regions of Interest Used in the Development and Application of Total Cloud Amount Diagnosis Methods in This Study.
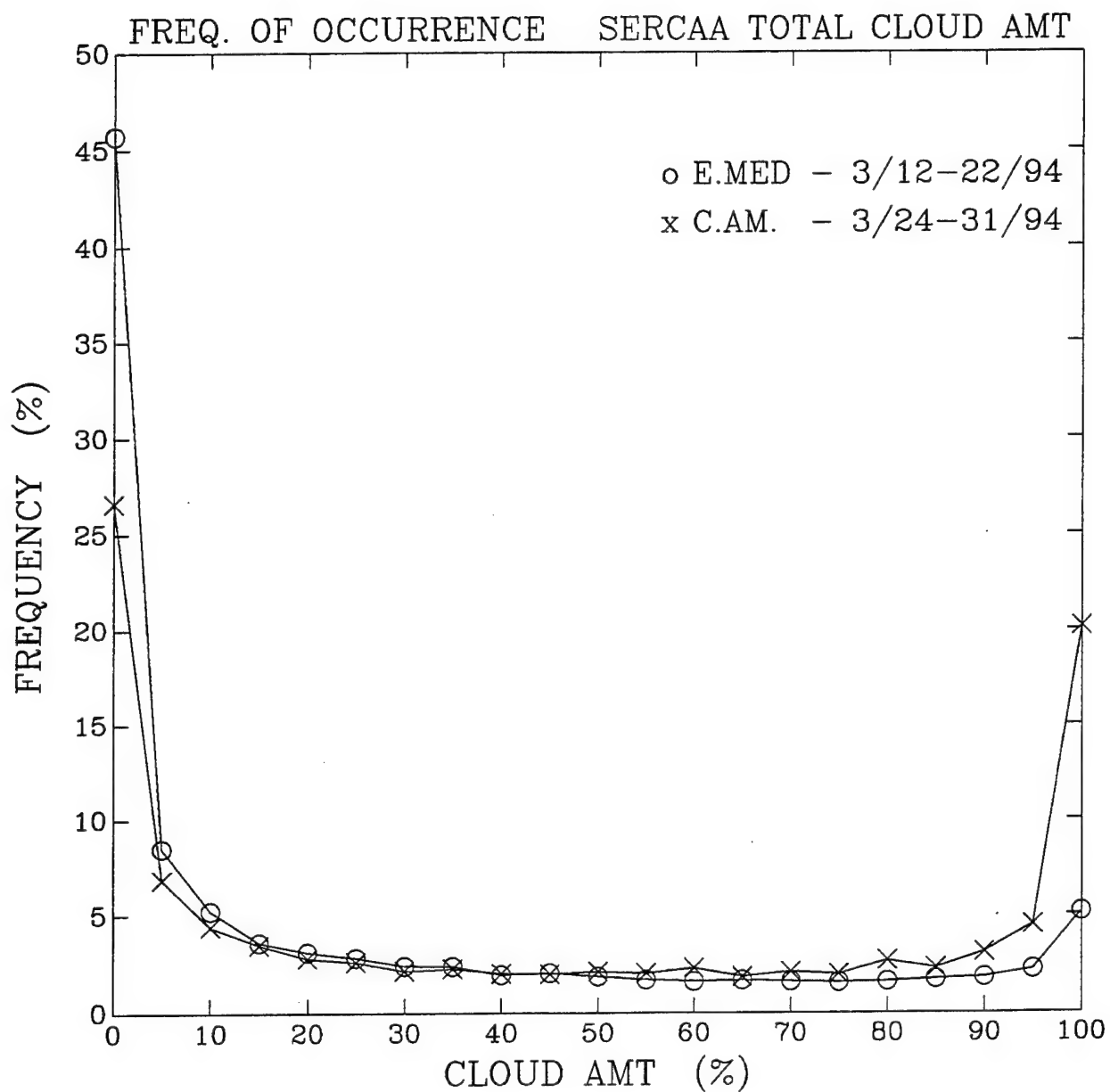
Figure 36. Frequency of Occurrence (Percentage of the Regional Transform Gridpoints for Four Times Per Day) of the Transformed SERCAA Total Cloud Amount.

quantities of wind components, temperature, specific humidity, surface pressure, and terrain height on our T106 model Gaussian grid. We then interpolated from 28 sigma layers to our 22 sigma layers, and spectrally transformed the results to T106 spherical harmonics. These were then used as initial conditions for twice-daily 48-hour PL-94 forecast executions, saving the forecasts at 6-hour intervals. The same predictor file generator as was described earlier in the report was used to generate predictor files at 00 and 12 UTC for 11-31 March. The predictors included are the same as those in the original list, as shown in Table 3.

We then executed the cloud diagnosis methods MLR (percent cloud cover to nearest 5 percent) and MDA (6-category cloud amount indices) as before, but now separately for the EMD and CNS regions. Over EMD, we used a 7-day development period for each of 4 days (7-day periods began on 12-15 March 1994), then applying the resulting predictand-predictor relationships to day eight forecasts in each case (forecasts initialized on 19-22 March 1994). Over CNS, we used two 5-day development periods (starting on 24-25 March 1994) and applied the resulting predictor-predictand relationships to forecasts initialized on 29 and 30 March 1994. In both regions, we performed development and application of relationships based on 12-hour forecasts only.

Because the regions and time periods for which SERCAA cloud cover analyses were available were limited, we had fewer equal-area gridpoints available in the development of the predictor-predictand relationships. In the 4 days of 7-day development periods for EMD, we had an average of 1521 gridpoints per development period available, and for the 2 days of 5-day development periods for CNS, we had an average of 1001 gridpoints available. These figures contrast with the availability of over 45,000 gridpoints available on average in the seven 10-day development periods when using the RTNEPH as predictand. As a consequence of the much smaller sample size, we found that the forward stepwise regression identified a development period average of just three predictors as contributing to the total correlation with cloud cover. However, even with such a limited number of predictors, we obtained total correlations of an average of 0.662 for EMD and 0.689

for CNS. Leading predictors were RH**2 and meridional wind for EMD, and RH and percent surface water cover for CNS.

As can be seen in Figure 36, the prevalence of zero cloud cover in the EMD region was over 45 percent. This is approaching the frequency of occurrence of zero cloud amount that we saw in RTNEPH high cloud in July (see Figure 2). As such, the MDA diagnosed high probabilities of zero cloud amount category. We found that 96 percent of the gridpoints in the independent sample were diagnosed as having a zero cloud cover by MDA, after the category selection was completed. In the CNS region only slightly more than 25 percent of the gridpoints were cloud-free in the 8- day total period of record. In this case, the MDA diagnosis resulted in just 27 percent cloud free points in the 2-day independent sample. The MDA also produced about 27 percent of its diagnosed gridpoints as category 6, which is 85-100 percent cloud cover.

We verified the 12-hour cloud cover predictions for forecasts initialized on 19-21 March 1994 over EMD and 29-30 March 1994 over CNS against the transformed SERCAA cloud cover analyses at the forecast valid times. This required the conversion of verifying SERCAA and MLR-diagnosed cloud cover amounts to 6-category cloud amount indices. Then categorical skill scores were generated for both MLR and MDA. These scores are shown for the two regions in Table 29.

Table 29. Comparison of Category Skill Scores for MLR and MDA from 12-Hour PL-94 Forecasts and SERCAA Total Cloud Cover for Four-Day (over EMD) and 2-day (CNS) Verification Periods.

	EMD		CNS	
Skill Score	MLR	MDA	MLR	MDA
Bias	0.143	-0.571	0.095	-0.352
RMSE	1.344	1.374	1.698	1.849
MAE	0.773	0.634	1.186	1.300
Percent Correct	54.7	70.1	36.3	34.1
Freq. of Occur. Fit	0.322	0.519	0.176	0.254

From the table, we see signs of the MDA's overestimate of the frequency of occurrence of zero cloud amount in EMD. We see a relatively high negative bias,

and an unsatisfactorily high frequency of occurrence fit value, reflecting the over-specification of gridpoints with zero cloud amount. Apparently, the category selection method was not able to overcome the high probability of clear in the sample. This suggests that the advantage of MDA over MLR in preserving the frequency distribution of the predictand is subject to sample size in the development of the predictor-predictand relationship. Another factor possibly influencing the performance of MDA may be the number (three) of predictors used to discriminate the categories. Inadequate sample size could have been a factor there also. In both regions, MLR produces values of bias and frequency of occurrence fit that are typical of the MLR values seen in the use of the more robust RTNEPH cloud amount diagnosis. This is not true for MDA--both scores are much larger (poorer) in the SERCAA regions. This suggests that the performance of MDA may be more sensitive to development sample size than the performance of MLR.

9. SUMMARY AND CONCLUSIONS

We attempted to establish a reliable method to deduce cloud cover predictions from non-cloud NWP forecasts. This required the development of a statistical relationship between hemispheric analyses of gridded cloud cover and gridded NWP model forecast fields for valid times of the cloud analyses. We derived such relationships for each forecast duration out to 48 hours and each of four vertical cloud decks (high, middle, low, and total cloud). The relationships were based on 10 days of twice-daily cloud analyses and NWP forecasts. The relationships were derived using the methods of multiple linear regression (MLR), MLR augmented by regression estimation of event probability (REEP), and multiple discriminant analyses (MDA). The separate relationships from all three methods were applied to an independent set of gridded forecast fields to deduce, or diagnose, the corresponding cloud cover spatial distributions. These were compared with the reference cloud analyses quantitatively and qualitatively to draw conclusions about the relative skill of the diagnostic methods.

133

We had observed in a previous study that MLR had the property of minimizing the root-mean-square error of the cloud amount forecast diagnoses. However, the MLR produced too many gridpoints with partial cloudiness and too few with clear or overcast compared with the reference cloud cover analysis. This motivated us to investigate what improvement might be gained by forcing the regression methodology to preserve the frequency of occurrence of clear and overcast in the reference cloud analysis. We did this by developing separate regressions for not-clear vs. clear points and not-overcast vs. overcast points using REEP. The resulting statistical relationships were used to discern which gridpoints in the application forecasts were likely clear and likely overcast, to preserve the correct number of each. All remaining points were diagnosed for cloud cover using the MLR.

We found that the hybrid REEP/MLR diagnosis method did indeed better preserve the frequency of occurrence of clear and overcast. This resulted in an increase in root-mean-square error relative to MLR as expected. It also had the effect of reducing the frequency of occurrence of nearly clear and nearly overcast conditions, which were already under-represented by MLR. Our conclusion was that we needed to be able to better preserve the frequency distribution of all categories of cloud cover rather than improve some at the expense of others.

This conclusion led to the development and testing of the method of MDA for diagnosis of cloud amount categories. We found that dividing the cloud coverage range of 0 to 100 percent into six categories results in the best compromise between loss of skill due to more categories and gain in skill relative to MLR in categorizing MLR cloud diagnoses. When both MDA-diagnosed cloud amount categories and categorized cloud cover diagnoses from MLR were compared to the reference cloud analyses rendered in categories, the MDA displayed much better agreement with the reference cloud analysis frequency distribution of categories. MDA also was consistently better than categorized MLR diagnoses in percent gridpoints correctly diagnosed, and was very competitive with MLR in root-mean-square error and mean absolute error.

Another benefit of the MDA method is that it produces an estimate of the probability of the diagnosed cloud amount category being correct. We found that this probability estimate yields useful information on the likelihood of a correct forecast of cloud cover. When MDA-diagnosed probabilities were high, very high category forecast accuracies were obtained.

Finally, because of the advantage that MDA has of accurately replicating the frequency distribution for the reference cloud analyses, we find that the MDA forecasts have more realistic spatial gradients of cloud cover. MLR produces too many gridpoints near 50 percent cloud amount, and this tends to reduce the sharpness of the cloud cover gradients.

From these results, we conclude that the MDA diagnosis method has promise, and should undergo further refinement and scrutiny as a cloud cover diagnostic technique. We feel that there may be room for improvement in cloud cover category diagnosis accuracy. This may be gained through more careful selection of forecast variables used to diagnose cloud cover, and through regional development of the statistical relationships.

In supplemental experiments, we tested the effect of (1) change of spatial resolution of the forecast fields, (2) use of supplementary forecast variables drawn directly from the forecast model, and (3) use of an alternative source of cloud cover reference. We found that the reduction of NWP forecast resolution and the addition of supplement forecast variables had no appreciable effect on cloud cover forecast accuracy. From this we conclude that limiting forecast predictors to fields of basic forecast variables at relatively coarse resolutions is not likely to appreciably compromise cloud diagnosis accuracy using the methods studied in this project. We also learned that the greatest impact of the alternative cloud reference was the limitation of sample size. Significant reductions in the size of the dependent data samples used to derive the forecast-cloud cover relationships are more likely to produce negative impacts on MDA than on MLR. From this we conclude that the performance of MDA is much more sensitive to developmental sample size than is the performance of MLR in diagnosing cloud cover from NWP forecast fields.

References

1. Norquist, D.C., Muench, HS., Aiken, D.L., and Hahn, D.C. (1994) *Diagnosing Cloudiness from Global Numerical Weather Prediction Model Forecasts*, PL-TR-94-2211, Phillips Laboratory, Hanscom AFB, MA, 139 pp. ADA289456

2. Kiess, R.B. and Cox, W. (1988) *The AFGWC Automated Real-Time Cloud Analysis Model*, AFGWC/TN-88-001, Air Force Global Weather Central, Offutt AFB, NE, 82 pp.

3. Hamill, T.M., d'Entremont, R.P., and Bunting, J.T. (1992) A Description of the Air Force Real-Time Nephanalysis Model, *Wea. Forecasting*, **7**: 288-306.

4. Crum, T.D. (1987) *AFGWC Cloud Forecast Models*, AFGWC/TN-87-001, Offutt AFB, NE, 66 pp.

5. Williamson, D.L. (1990) Semi-Lagrangian Moisture Transport in the NMC Spectral Model, *Tellus*, **42A:** 413-428.

6. Schiffer, R.A. and Rossow, W.B. (1985) ISCCP Global Radiance Data Set: A New Resource for Climate Research, *Bull. Amer. Meteor. Soc.*, **66**: 1498-1505.

7. Rossow, W.B. and Garder, L.C. (1993) Validation of ISCCP Cloud Detections, *J. Climate*, **6**: 2370-2393.

8. Hortal, M. (1990) Use of Reduced Gaussian Grids in Spectral Models, *ECMWF Newsletter*, Number 52, December 1990, European Center for Medium-Range Weather Forecasts, 3-7.

9. Norquist, D.C., Yang, C., Chang, S., and Hahn, D.C. (1992) *Phillips Laboratory Global Spectral Numerical Weather Prediction Model*, PL-TR-92-2225, Phillips Laboratory, Hanscom AFB, MA, 154 pp. ADA267239.

10. Norquist, D.C. and Chang, S.S. (1994) Diagnosis and Correction of Systematic Humidity Error in a Global Numerical Weather Prediction Model, *Mon. Wea. Rev.*, **122**: 2442-2460.

11. Yang, C., Hahn, D.C., Chang, S., and Aiken, D.L. (1995) *Modifications to the Representation of the Surface Layer Processes in the Phillips Laboratory Global Spectral Model,* PL-TR-95-2111, Phillips Laboratory, Hanscom AFB, MA, 70 pp.ADA310614

12. Miller, M.J., Beljaars, A.C.M., and Palmer, T.N. (1991) The Sensitivity of the ECMWF Model to the Parameterization of Evaporation from Tropical Oceans, *J. Climate,* **5:** 418-434.

13. Jaeger, L. (1983) Monthly and Areal Patterns of Mean Global Precipitation, In *Variations in the Global Water Budget,* A. Street-Perrott, M. Beran, and R. Ratcliffe, Eds., D. Reidel Publishing Co., Dordrecht, Holland, 129-140.

14. Peixoto, J.P. and Oort, A.H. (1983): The Atmospheric Branch of the Hydrological Cycle and Climate, In *Variations in the Global Water Budget,* A. Street-Perrott, M. Beran, and R. Ratcliffe, Eds., D. Reidel Publishing Co., Dordrecht, Holland, 5-65.

15. Schattel, J. (1992) *Refinement and Testing of the Radiative Transfer Parameterization in the PL Global Spectral Model,* PL-TR-92-2169, Phillips Laboratory, Hanscom AFB, MA, 27 pp. ADA256840.

16. Slingo, J.M. (1987) The Development and Verification of a Cloud Prediction Scheme for the ECMWF Model, *Quart. J. Roy. Meteor. Soc.,* **113:** 899-927.

17. Mahrt, L., Ek, M., Kim, J., and Holtslag, A.A.M. (1991) *Boundary Layer Parameterization for a Global Spectral Model,* PL-TR-91-2031, Phillips Laboratory, Hanscom AFB, MA, 120 pp. ADA235310.

18. Miller, R.G. (1964) *Regression Estimation of Event Probabilities,* Travelers Research Center, Hartford, Contract CWB-10704, Tech. Rep. No. 1, 153 pp.

19. Glahn, H.R., Murphy, A.H., Wilson, L.J., and Jensenius, J.S. (1991) Statistical Models - Regression, In *Lectures Presented at the WMO Training Workshop on the Interpretation of NWP Products in Terms of Local Weather Phenomena and Their Verification,* Wageningen, The Netherlands, 29 July-9 August 1991, PSMP Report Series No. 34, WMO/TD, No. 421, World Meteorological Organization, v1-v22.

20. Glahn, H.R., Murphy, A.H., Wilson, L.J., and Jensenius, J.S. (1991) Statistical Models - Discriminant Analysis, In *Lectures Presented at the WMO Training Workshop on the Interpretation of NWP Products in Terms of Local Weather Phenomena and Their Verification,* Wageningen, The Netherlands, 29 July-9 August 1991, PSMP Report Series No. 34, WMO/TD, No. 421, World Meteorological Organization, vi1-vi19.

21. Miller, R.G. (1962) Statistical Prediction by Discriminant Analysis, *Meteorological Monograph,* Vol. 4, No. 25, Amer. Meteor. Soc., Boston, MA, 53 pp.

22. Allen, G. and LeMarshall, J.F. (1994) An Evaluation of Neural Networks and Discriminant Analysis Methods for Application in Operational Rain Forecasting, *Aust. Met. Mag.,* **43:** 17-28.

23. Gustafson, G.B., Isaacs, R.G., d'Entremont, R.P., Sparrow, J.M., Hamill, T.M., Grassotti C., Johnson, D.W., Sarkisian, C.P., Peduzzi, D.C., Pearson, B.T., Jakabhazy, V.D., Belfiori, J.S., and Lisa, A.S. (1994) *Support of Environmental Requirements for Cloud Analysis and Archive (SERCAA),* PL-TR-94-2114, Phillips Laboratory, Hanscom AFB, MA, 105 pp. ADA283240.

24. Walcek, C.J. (1994) Cloud Cover and Its Relationship to Relative Humidity During a Springtime Midlatitude Cyclone, *Mon. Wea. Rev.,* **122:** 1021-1035